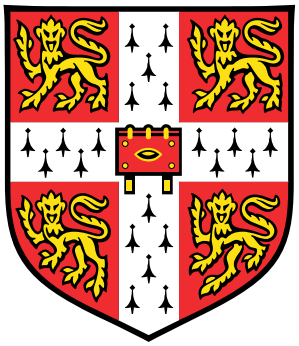
ADVANCED MAGNETIC RESONANCE IMAGING OF OSTEOARTHRITIS

Dr James William MacKay



This dissertation is submitted for the degree of Doctor of Philosophy

Magdalene College

July 2019

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the Degree Committee for Clinical Medicine & Veterinary Medicine.

ABSTRACT

This thesis examines the potential utility of magnetic resonance (MR) quantitative imaging biomarkers (QIBs) of knee osteoarthritis (OA) for rapid assessment of treatment efficacy in experimental medicine studies.

The development of treatments able to modify disease in OA is hampered by an inability to evaluate treatment response over a timeframe relevant to clinical trials. There are particular challenges in the experimental medicine setting due to the small numbers of participants and short follow-up duration relative to the expected time course of OA development and progression. Multiple MR QIBs of OA exist which may help address the problem of early evaluation of treatment response. However, their use in early phase studies has remained limited. Possible reasons for this include incomplete characterisation of the performance of QIBs in this setting and lack of head-to-head comparison of candidate QIBs to determine which would be optimal.

This thesis aims to address these shortcomings and provide new information on the likely utility of MR QIBs in the setting of experimental medicine studies, as well as their potential for improving our general understanding of OA pathophysiology.

I start by examining the reliability and ability to discriminate between OA and healthy knees of cartilage compositional MR imaging in a systematic review and meta-analysis. I then describe the development and validation of a novel semi-automatic surface-based method for analysing articular cartilage composition and morphology at the knee which may offer improved responsiveness and spatial localisation of change. Moving to QIBs of subchondral bone, I evaluate the association between measures of subchondral bone architecture derived from MR texture analysis and OA progression in the Osteoarthritis Initiative. The remainder of the thesis describes a prospective observational study where the utility of MR QIBs of synovium, subchondral bone and cartilage in experimental medicine studies is assessed.

In summary, this thesis will inform decisions regarding the use of MR-based QIBs in future longitudinal and interventional studies. Their inclusion in experimental medicine studies may allow early assessment of treatment efficacy at a structural level and improve efficiency of treatment development pipelines.

ACKNOWLEDGEMENTS

My 3 and a bit years of PhD research have been the most challenging but also the most enjoyable and rewarding of my career to date. There are many people who are responsible for making this the case, too many to acknowledge individually. However, there are some deserving of special mention.

I consider myself extremely lucky to have undertaken my PhD under the supervision of Professor Fiona Gilbert. As well as guiding my research, she has taught me a lot about how to develop a successful research career and has provided an example of what outstanding research leadership looks like. The co-supervision of Professor Andrew McCaskie has been invaluable, particularly the key clinical insights that he has provided into the challenges of treating osteoarthritis. His vision for musculoskeletal research in Cambridge is inspiring and he has been a pleasure to work with.

Professor Andoni Toms in Norwich has been an excellent mentor throughout my research career to date. I owe him a debt of gratitude for starting me on my radiological research journey as a young registrar. It has been great to be able to bounce ideas off someone external to my project during my PhD, and I can always count on him to spot at least one grammatical error or formatting *faux pas* in anything I send him to review.

My PhD was funded by the Addenbrooke's Charitable Trust and the Experimental Medicine Initiative (EMI). I am grateful to both parties for enabling me to undertake this work. Through the EMI, I have had the pleasure of meeting and working with many excellent people within GlaxoSmithKline including Rob Janiczek, Alex Morgan-Roberts, Jon Larkin, Danielle Gerlag and Duncan Richards. They have helped me to navigate the industry/academia border and I have learned a lot by doing so.

I was fortunate to have access to additional advice and guidance through the EMI from Professor Ian Wilkinson, and I thank him for his support. I would also like to acknowledge the assistance of Nicola White, the EMI programme manager, who has been very helpful throughout.

The dynamic contrast enhanced (DCE) MR work described in this thesis was performed in collaboration with BiOxyDyn. I would like to thank Faezeh Sanaei and Jo

Naish for their hard work on the data analysis (and for answering my many questions), and to Prof Geoff Parker for his valuable input.

I am fortunate to now be able to call many of my colleagues within the Department of Radiology friends, and I hope to keep in touch on both an intellectual and a personal level in years to come. Special mention must go to Josh Kaggie and Martin Graves, whose MR physics expertise has been invaluable and whose general good humour means that they are always a pleasure to be around. Dimitri Kessler arrived as a new PhD student towards the end of my PhD and it has been a pleasure helping him to embark upon his journey. The departmental administrative staff have also been wonderful. In particular, Candice Anderson and Sarah Perkins have shown incredible patience and understanding in dealing with my many questions and requests and have helped to keep me organised. I will miss the friendly and welcoming environment which they have created.

Research is at its best as a collaborative endeavour, and I have been lucky to collaborate with some excellent people during my PhD. Special thanks are reserved for Stephen McDonnell and Wasim Khan in the Division of Trauma & Orthopaedics and Graham Treece and Tom Turmezei in the Department of Engineering. Simon White of the MRC Biostatistics Unit was also very helpful in providing initial guidance to my prospective study.

On that note, I am incredibly grateful to the participants in this prospective study who willingly gave up their time to lie in our MR scanner and have their knees imaged on multiple occasions. One of the most rewarding parts of my PhD was getting to know them and understanding how their OA was affecting their lives. Their frustration at the lack of effective treatment options will continue to be a real motivator for me in my future work.

Finally, the support of my friends and family has been unwavering throughout this PhD.

My parents provided me with an outstanding education and encouraged me to ask questions from an early age, something which has no doubt led me to embrace a career in research. I am very grateful to them for this.

Hazel was born about two-thirds of the way into this PhD. While her arrival and the subsequent months have felt something like being at the centre of a whirlwind, she has brought untold joy to my life. At times during my PhD it has felt like the work is all-encompassing but one little smile from her and I am reminded of what is truly important.

Eilbhe has been my rock and teammate since long before embarking on my PhD. Everything I have achieved and will achieve is a credit to her support and love.

PUBLICATIONS ARISING FROM THIS THESIS

MacKay JW, Low SBL, Smith TO, Toms AP, McCaskie AW, Gilbert FJ. Systematic review and meta-analysis of the reliability and discriminative validity of cartilage compositional MRI in knee osteoarthritis. Osteoarthr Cartil. 2018;26(9):1140–1152.

MacKay JW, Kapoor G, Driban JB, et al. Association of subchondral bone texture on magnetic resonance imaging with radiographic knee osteoarthritis progression: data from the Osteoarthritis Initiative Bone Ancillary Study. Eur Radiol. 2018;28(11):4687–4695.

PRESENTATIONS ARISING FROM THIS THESIS

Oral

2019

'Utility of MRI quantitative biomarkers in experimental medicine studies of knee OA'. International Workshop on Osteoarthritis Imaging (IWOAI), Prince Edward Island, Canada.

'Cartilage compositional MRI: does it have a role in defining early OA' [invited]. Osteoarthritis Research Society International annual congress (OARSI), Toronto, Canada.

2018

'Physiology of joint tissues in osteoarthritis' [invited]. International Society for Magnetic Resonance in Medicine annual meeting (ISMRM), Paris, France.

2017

'Association between MRI subchondral bone texture and radiographic OA progression: Data from the Osteoarthritis Initiative Bone Ancillary Study'. IWOAI, Sydney, Australia

'High resolution 3D modelling of cartilage surfaces at the knee using clinical magnetic resonance imaging data'. IWOAI, Sydney, Australia.

Poster

2019

'Cartilage imaging for experimental medicine studies: 6-month changes in morphology and composition can be detected by 3D cartilage surface mapping of knee magnetic resonance imaging data'. OARSI, Toronto, Canada.

2018

'Repeatability and discrimination validity of cartilage imaging biomarkers for experimental medicine studies of knee OA'. OARSI, Liverpool, UK.

*'Dynamic contrast enhanced MR imaging in early stage knee osteoarthritis: A test-retest repeatability study'. ISMRM, Paris, France.

2017

'Systematic review of the reliability and discriminative validity of cartilage compositional MRI in knee osteoarthritis'. IWOAI, Sydney, Australia.

**'3D modelling of cartilage surfaces at the knee using clinical magnetic resonance imaging data' . ISMRM, Hawaii, USA.

* First author & presenter: Dr Faezeh Sanaei, BiOxyDyn, Manchester, UK. I was second author, co-wrote the abstract and critically reviewed the final poster design.

** Presented by Dr Josh Kaggie (second author) as I was unable to attend the conference.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

COMMONLY USED ABBREVIATIONS

2D	2-dimensional
3D	3-dimensional
3D-CaSM	3D Cartilage Surface Mapping
3T	3 Tesla
AMROA	Advanced MR imaging of osteoarthritis
ANCOVA	Analysis of covariance
AUC	Area under the receiver operating characteristic curve
BML	Bone marrow lesion
CI	Confidence interval
DCE-MR	Dynamic contrast enhanced magnetic resonance (imaging)
dGEMRIC	Delayed gadolinium enhanced MR imaging of cartilage
DICOM	Digital Imaging and Communications in Medicine
DMOAD	Disease modifying OA drug
DMOAT	Disease modifying OA treatment
EM	Experimental medicine
FDA	Food and Drug Administration
FNIH	Foundation of the National Institutes for Health
FSA	Fractal signature analysis
FSE	Fast spin echo
gdCA	Gadolinium based contrast agent
HRpQCT	High resolution peripheral quantitative computed tomography
HV	Healthy volunteer
$IAUC_{60}$	Area under the time-concentration curve at 60s
ICC	Intraclass correlation coefficient
IQR	Interquartile range
IW	Intermediate-weighted
JSN	Joint space narrowing
JSW	Joint space width
KL	Kellgren-Lawrence grade
KOOS	Knee injury and osteoarthritis outcome score
K^{trans}	Volume transfer coefficient between the plasma and the extracellular extravascular space
LASSO	Least absolute shrinkage and selection operator (regression)
MOAKS	MRI Osteoarthritis Knee Score
MR(I)	Magnetic resonance (imaging)

OA	Osteoarthritis
OAI	Osteoarthritis Initiative
OARSI	Osteoarthritis Research Society International
OMERACT	Outcome measures in rheumatology
QIB	Quantitative imaging biomarker
RMSCV	Root-mean-square average coefficient of variation
ROI	Region of interest
SBA	Subchondral bone area
SD	Standard deviation
SDD	Smallest detectable difference
SMD	Standardised mean difference
SPGR	Spoiled gradient echo
SPM	Statistical parametric mapping
STV	Synovial tissue volume
T1	MR longitudinal relaxation time constant
T1rho	MR longitudinal relaxation time constant in the presence of a radiofrequency field
T2	MR transverse relaxation time constant
TA	Texture analysis
v_e	Proportion of extracellular extravascular space
VEP	Volume of enhancing pannus
v_p	Proportion of blood plasma volume

KEY COLLABORATOR CONTRIBUTIONS

Chapter 2 – Systematic review and meta-analysis…

Dr Samantha Low performed initial screening to identify potentially eligible studies per inclusion and exclusion criteria in duplicate with me. Dr Low also double-checked the data extraction for included studies.

Dr Toby Smith and Prof Andoni Toms provided important systematic review expertise to guide my efforts.

Chapter 3 – 3D cartilage surface mapping validation…

Dr Graham Treece and Dr Tom Turmezei provided training in the software used for 3D cartilage surface mapping, helped to design the validation experiment and provided critical feedback on the results.

Chapter 4 – Association of subchondral bone texture…

Dr Geeta Kapoor performed duplicate texture analysis for inter-observer reproducibility assessment.

Chapter 6 – Dynamic contrast enhanced…

I performed the image analysis jointly with Dr Faezeh Sanaei and Dr Jo Naish at BiOxyDyn. Specifically, I created the manual regions of interest which were then incorporated into the BiOxyDyn DCE analysis pipeline. I reviewed the resulting parameter maps and performed subsequent data analysis using the parameters outputted from the BiOxyDyn pipeline.

Chapter 8 – 3D cartilage surface mapping implementation…

Dr Graham Treece and Dr Tom Turmezei provided training in the software used for 3D cartilage surface mapping and provided critical feedback on the results.

Dr Tom Turmezei performed duplicate 3D-CaSM and manual segmentation for inter-observer reproducibility assessment.

CHAPTER 1:

INTRODUCTION

1.1 KNEE OSTEOARTHRITIS

1.1.1 Epidemiology & current treatment options

Osteoarthritis (OA) is the commonest cause of pain and loss of function in the adult population, affecting an estimated 8.75 million individuals in the UK(1). Although traditionally considered a disease of the elderly, people of working age (under 65 years old) make up 2/3 of this total, and in 2016 (the latest year for which data are available) OA was the commonest reason for claiming disability living allowance in the UK(1,2). The prevalence of OA is expected to increase over the next few decades given the rising ageing population and increasing prevalence of obesity which are both important risk factors for developing OA(3). The commonest joint involved is the knee, which is the focus of this thesis(1).

The term 'osteoarthritis' refers to a clinical syndrome of joint pain associated with various structural abnormalities. Knee OA (and indeed OA in general) is a heterogeneous condition with distinct OA phenotypes increasingly being described(4). These involve variations in underlying aetiology (e.g. post-traumatic OA), joint structures involved (e.g. atrophic vs hypertrophic OA) and natural history (e.g. rapidly progressive OA). The conventional view of knee OA as primarily a cartilage disease has been thoroughly debunked, and it is now recognised to be a disease of multiple joint structures including bone, synovium, ligaments, menisci, capsule and juxta-articular muscle as well as cartilage(5).

Current treatment options for knee OA depend on stage of disease. Total knee arthroplasty (TKA) and unicompartmental knee arthroplasty (UKA) are effective treatments for end-stage knee OA(6,7). However, these options are less suitable earlier in the disease process due to limited lifespan of the prostheses and associated functional limitations making joint replacement less acceptable to younger patients wishing to maintain relatively high degrees of physical activity. The mainstay of treatment for individuals with earlier disease is non-pharmacological, including lifestyle modification (particularly weight loss) and physiotherapy. Previously, knee arthroscopy with or without partial meniscectomy was commonly performed in patients with earlier stages of OA. However, recent randomised controlled trial evidence has suggested that arthroscopy offers little benefit over conservative

treatment although it should be noted that the generalisability of these results is disputed(8–10). Recent guidelines from the British Association for Surgery of the Knee (BASK) have aimed to clarify situations in which knee arthroscopy may be an appropriate treatment in the presence of OA. These guidelines introduced the concept of the 'meniscal target' lesion, based on the presence or absence of displaced meniscal tear or displaced meniscal tissue. In patients with mild-to-moderate knee OA, BASK guidelines suggest that arthroscopy may be appropriate in the presence of a target lesion and following an appropriate period of conservative treatment(11).

Pharmacological options in early OA are limited at present to symptomatic relief with oral analgesics (which are often accompanied by undesirable side effects) and short-acting intra-articular treatments (mainly corticosteroids). At present, there is no treatment for early disease that has conclusively demonstrated disease modifying properties, i.e. the ability to prevent or reverse progression of OA. There is therefore a large group of individuals experiencing debilitating symptoms but not yet suitable for knee replacement surgery for whom no effective pharmacological treatment exists(12).

1.1.2 Pathophysiology of OA

Pathological changes in knee OA include degeneration of articular cartilage and meniscus, alterations in subchondral bone, synovial inflammation, ligamentous degradation and changes in periarticular muscles(13). The typical description of OA as a disease of 'wear and tear' is now recognised to be outdated, with current concepts viewing OA onset as being caused by an imbalance between normal wear and repair processes(14).

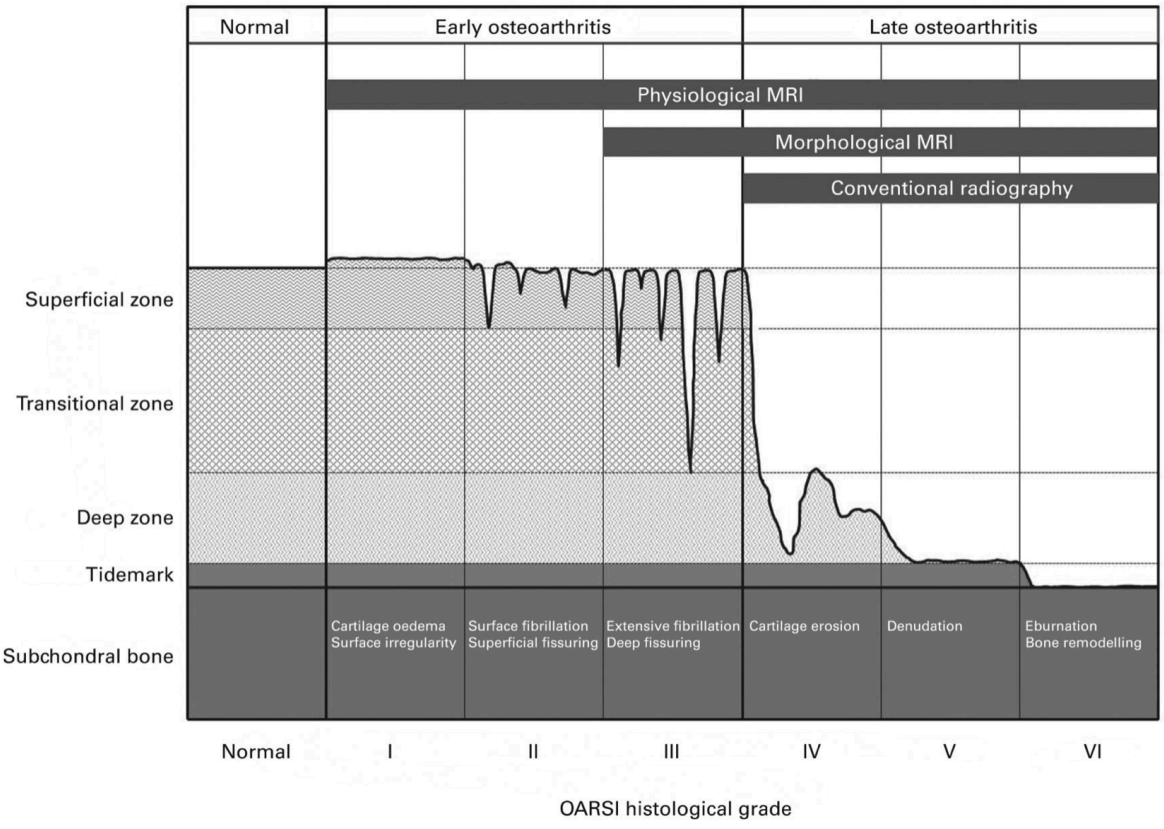
Normal articular cartilage exists to provide a surface for smooth, friction-free movement. It acts in concert with the subchondral bone to transmit loads applied to the joint. The superficial layers secrete lubricin, a lubricating component of the synovial fluid. The only cell type found in healthy articular cartilage is the chondrocyte which plays a crucial role in cartilage homeostasis(15).

In OA, cartilage alterations at the microscopic and molecular level include a phenotypic shift with chondrocyte clustering and hypertrophy, increased catabolic activity with proteases causing degradation of the cartilage matrix, and eventual chondrocyte senescence or apoptosis. This proceeds as a 'vicious circle' with

senescent/apoptotic chondrocytes stimulating further chondrocyte activation, secretion of pro-inflammatory and catabolic mediators and further matrix degradation(16). These microscopic level changes directly link to the macroscopic changes that are seen on imaging (Figure 1.1). An initial increase in protease activity leads to a reduction in proteoglycan concentration. This reduces the amount of bound water within the cartilage and causes a net increase in water mobility, reducing the cartilage's load bearing capacity. As matrix degradation progresses, damage to the collagen lattice allows further increase in water mobility with associated reduction in load-bearing function(17). At this point, the damage is already irreversible as there is minimal cartilage collagen synthesis in humans after adolescence [HEINEMEIER 2016].

Figure 1.1

Illustration of the changes in cartilage morphology occurring with increasing degradation. The theoretical sensitivity of different imaging modalities to detect each stage of degradation is displayed as dark grey bars at the top of the figure. Reproduced from (17).



The subchondral bone forms a biomechanical couple with the overlying articular cartilage. Its main function is to absorb and respond to load, as well as providing

nutrients to the deeper layers of articular cartilage. Normal subchondral bone homeostasis consists of a balance between bone formation and bone resorption. Each of these processes is associated with a specific cell type, the osteoblast (formation) and osteoclast (resorption). In response to a stimulus (e.g. increased or reduced loading), the subchondral bone trabecular surfaces are 'activated'. This activation involves both osteoclasts and osteoblasts. Osteoblasts deposit osteoid which is mineralised via deposition of calcium hydroxyapatite to form mature trabecular bone. Osteoclasts then selectively resorb this, followed by further osteoblast-mediated bone deposition(18). This coupled osteoblast-osteoclast activity allows tightly controlled remodelling of trabecular network, permitting the bone to adapt to its local loading conditions per Wolff's law(19).

Co-ordination of bone formation and bone resorption is lost in OA. Initially there is a state of increased bone turnover with an increase in both processes. However, because the bone is being formed at a supraphysiological rate, mineralisation cannot keep up and the new bone formed has inferior material properties. This explains the observation that subchondral bone volume fraction increases, but with paradoxically reduced stiffness(20). It is unclear what the initiating event for this state of increased bone turnover is. Proposed aetiologies include biomechanical (increased loading and/or altered loading pattern) and biological (local or systemic inflammation) factors. Reduced load-bearing capacity of the subchondral bone will have a direct knock-on effect on the load-bearing overlying articular cartilage, increasing the shear and strain forces that it experiences(18). As well as the adverse mechanical effects on articular cartilage, vascular invasion through the osteochondral junction causes dysregulation of normal bone-cartilage molecular 'crosstalk' and promotes cartilage breakdown(21). Concurrent or closely sequential changes in subchondral bone and overlying articular cartilage have been observed in animal models(22).

Closely linked to the described changes in bone and cartilage are alterations in the synovium. Normal synovium is a thin layer of connective tissue which seals the joint cavity from adjacent tissues. In health, the major cell type is the synovial fibroblast. Its principle function is to secrete and regulate the composition of synovial fluid. Synovial fluid lubricates the joint via molecules such as lubricin and hyaluronic acid, and is the

major source of nutrients to the superficial layers of articular cartilage (deep layers also receive nutrition from the subchondral bone)(23).

Synovial changes in OA are characterised by an influx of leucocytes (particularly macrophages and T-lymphocytes), hypertrophy and hyperplasia of the synovial lining layer and angiogenesis(24). The triggering event for these changes is debated, but likely relate to the production of an increased amount of joint debris (e.g. due to increased cartilage breakdown), or an abnormal response to normal amounts of joint debris due to 'priming' of synovial immune cells by local or systemic factors. Inflamed synovium produces both proteolytic enzymes and pro-inflammatory mediators (e.g. cytokines) which promote the expression of proteolytic enzymes within chondrocytes, leading to cartilage damage(25). In addition, activated synovial macrophages have been shown to play a key role in osteophyte formation in animal models(26).

1.2 IMAGING OF KNEE OSTEOARTHRITIS

1.2.1 Plain radiography

Plain radiographs have traditionally formed the basis of radiological diagnosis and classification of knee OA. These have the advantage of being cheap and quick to perform. Although there is an associated ionising radiation dose, this is minimal (< 1 μSv, less than one day of 'background' radiation in the UK). Classification of OA on plain radiographs is most often performed using the Kellgren-Lawrence (KL) system(27). This categorises radiographs into normal, doubtful OA, mild OA, moderate OA, and severe OA classes based on the presence of osteophytes, joint space narrowing, and subchondral bone attrition (Table 1.1).

Table 1.1

Kellgren-Lawrence grading of osteoarthritis

Grade	Description
0	Normal
1	Doubtful - Possible osteophytic lipping, minor joint space narrowing (JSN)
2	Mild – Definite osteophyte, little to no JSN
3	Moderate – Definite JSN + osteophytes, early subchondral sclerosis +/- early subchondral bone deformity
4	Severe – Severe JSN, large osteophytes, severe sclerosis + definite bone deformity

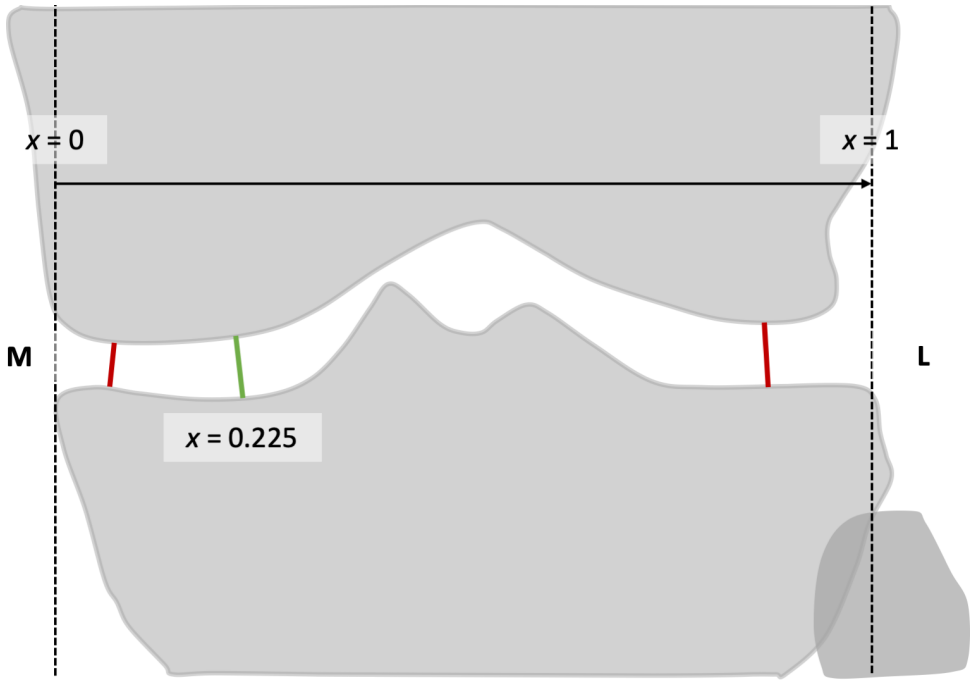
The Osteoarthritis Research Society International (OARSI) grading system is a modification of the KL system (28). This scores joint space narrowing, osteophytes and subchondral sclerosis separately and allocates distinct scores for the medial and lateral compartments. Qualitative grading systems such as KL and OARSI have the advantage of being quick to perform and easy to interpret. However, inter-observer agreement is suboptimal and neither system is responsive to change(29). The KL system in particular has the shortcoming of assuming a linear progression of disease involving development of osteophytes prior to joint space narrowing. In fact, a large number of patients will develop joint space narrowing prior to radiographic osteophyte development. Such patients do not fit neatly into the KL classification system(30).

Quantitative methods of radiographic assessment have also been developed. Measurement of tibiofemoral joint space width (JSW) is probably the best established of these. This can be performed using manual measurements, but the advent of automated techniques has improved the reliability of JSW measurement and has enabled the rapid assessment of JSW in both large epidemiological studies and clinical trials(31). Until recently, JSW measurement was the only approved imaging metric accepted by the United States (US) Food and Drug Administration (FDA) for use in clinical trials(32). There are two main types of JSW measurement used by researchers. Most widely used is minimum JSW (minJSW), which reflects the smallest distance between the femoral and tibial subchondral bone for a given tibiofemoral compartment (i.e. medial or lateral). Fixed location measurements of JSW are also used, which are measurements taken at a consistent location in the medial to lateral dimension normalised to the size of the knee using a co-ordinate system as

demonstrated in Figure 1.2(33). Fixed location JSW has demonstrated superior responsiveness to minJSW although both measures are still commonly used(34). Measurement of JSW has the advantages inherent to plain radiographic methods of low cost and accessibility, with the additional benefit over qualitative techniques of improved sensitivity to change and improved reliability with automated methods. However, JSW measurement also has several disadvantages. First, differences in knee positioning between individuals or within the same individual over time can confound measurement. To some extent this can be overcome using standardised positioning protocols and assistive devices(35). Second, JSW is a composite measure reflecting both cartilage thickness and meniscal position(36,37). This lack of specificity can be troublesome, for example in determining whether a change in JSW is due to change in cartilage thickness or meniscal extrusion in a trial of a potential disease modifying OA treatment (DMOAT) targeting cartilage. Finally, JSW is relatively unresponsive when compared to magnetic resonance (MR) imaging-based measures, discussed in the next section(38).

Figure 1.2

Schematic illustration of a posteroanterior knee radiograph. Red line indicates location of minJSW in the medial (M) and lateral (L) compartments. Dotted vertical black lines indicate boundaries of medial and lateral joint space used to derive co-ordinate system for fixed location JSW. Green line indicates measurement of fixed location JSW at $x = 0.225$, i.e. 22.5% of the distance between the medial and lateral vertical dotted lines.



Additional quantitative measures can also be derived from radiographs, for example fractal signature analysis (FSA) of subchondral bone. FSA quantifies the degree of self-similarity of an image and provides an indirect measure of trabecular microstructure. This has demonstrated reasonable construct and predictive validity, and does not suffer from the same sensitivity to positioning as does JSW(39–41). However, the output of FSA is complex and the biological meaning of the parameters derived from FSA is not always clear.

1.2.2 Magnetic resonance imaging

MR imaging has the substantial advantage over plain radiography of being able to directly visualise all joint tissues involved in the OA disease process, including subchondral bone, articular cartilage, ligament, meniscus, and synovium(13). Moreover, it has the ability to detect earlier stages of the disease before radiographic changes are apparent(42). Despite this, the use of MR imaging in routine clinical practice for OA diagnosis and monitoring is not recommended due to the relatively high cost, limited influence on current management decisions and potential association with higher rates of arthroscopic intervention(43).

However, in the research setting MR imaging is widely used and recommended(44). Two general approaches to MR-based assessment of OA-related abnormalities exist: semiquantitative and quantitative. Semi-quantitative grading systems typically divide the knee into anatomical subregions then grade various pathologies (e.g. cartilage defects, bone marrow lesions – BML) in each subregion according to an ordinal scale of severity (Table 1.2).

Table 1.2

Summary of the most commonly used MR semi-quantitative scoring methods for knee OA. Adapted from (45).

Scoring system	Year of publication	Scored features (grades)
WORMS	2004	Cartilage (0–6); BMLs (0–3); subchondral cysts (0–3); bone attrition (0–3); effusion and synovitis (0–3); periarticular cysts (0–3); bursitides (0–3); loose bodies (0–3); osteophytes (0–7); meniscal tear (0–4); cruciate and collateral ligaments (0–1)
BLOKS	2008	Cartilage size and depth (0–3, plus extent of any cartilage loss at specified point); BMLs (0–3, for each lesion); osteophytes (0–3); effusion (0–3); meniscal extrusion (0–3); synovitis (in Hoffa's fat pad 0–3 and at 5 additional sites 0–1); meniscal status (0–1 for intrameniscal signal, tears, maceration, meniscal cyst, each scored individually); ligaments (0–1); periarticular cysts/bursitis (0–1); loose bodies (0–1)
MOAKS	2011	Cartilage size and depth (0–3); BMLs (0–3, for each lesion); osteophytes (0–3); effusion-synovitis (0–3); Hoffa synovitis (0–3); meniscal extrusion (0–3); meniscal status (0–1, for intrameniscal signal, tears, maceration, meniscal cyst, hypertrophy; scored individually); ligaments (0–1); periarticular cysts/bursitides (0–1, scored individually); loose bodies (0–1)

Abbreviations: WORMS – Whole Organ MRI score, BLOKS – Boston-Leeds Osteoarthritis of the Knee Score, MOAKS – MRI Osteoarthritis Knee Score

The most widely used scoring method at present is the MOAKS (MRI Osteoarthritis Knee Score) system(46). This was designed to harmonise existing scoring methods, particularly BLOKS (Boston-Leeds Osteoarthritis of the Knee Score) and WORMS (Whole Organ MRI score), while addressing some of their shortcomings(47,48). Many of the semiquantitative MR features graded by MOAKS have a stronger association with symptoms than do plain radiographic measures(49,50). This is particularly true for synovitis and BML. Moreover, semiquantitative MR features have improved responsiveness to change than do radiographic measures(51–53). Semiquantitative scoring systems typically use standard clinical MR pulse sequences and so are widely applicable, illustrated by their successful implementation in several large observational and interventional studies(54,55). However, there are limitations associated with their use. Most important, they are less responsive than quantitative MR measures for assessing comparable features(51,56). For example, quantitative cartilage thickness

measurements are more responsive than semiquantitative cartilage scores. Second, they are inherently subjective in nature and therefore subject to inter-observer error(46). Finally, the ordinal nature of the data plus the multiple different features scored across multiple subregions makes statistical analysis complex both to perform and interpret(57).

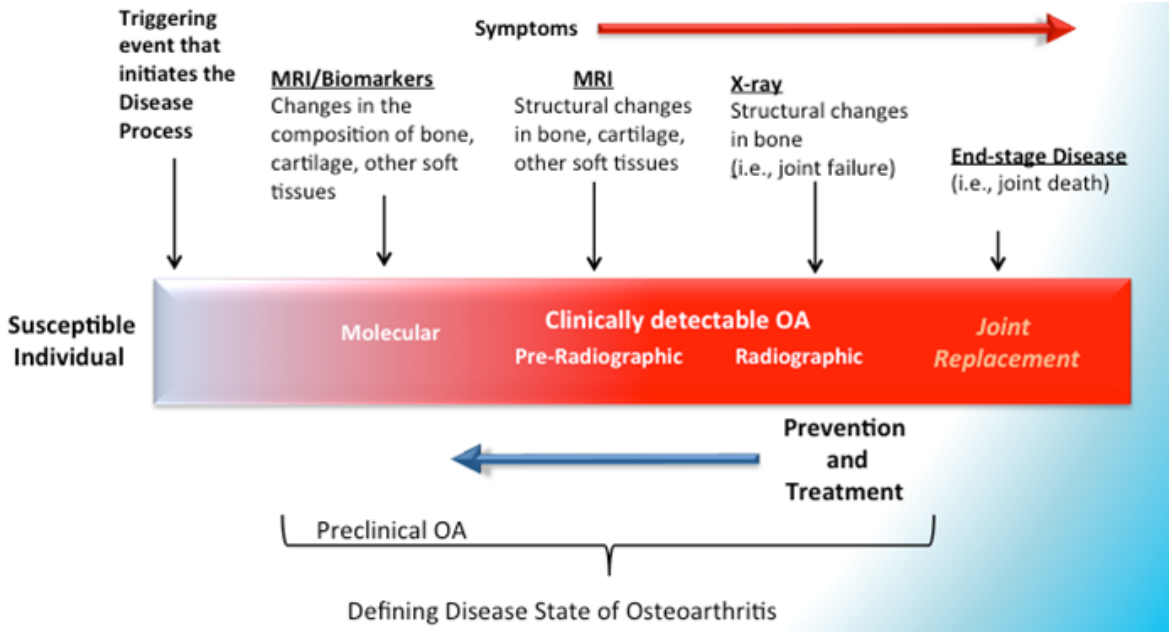
Quantitative approaches to MR imaging can overcome some of these shortcomings. Such approaches can be broadly divided into morphological and physiological categories. Morphological approaches involve quantification of the volume, area, thickness or position of various joint tissues. One of the best-established of these measures is quantification of cartilage thickness. This is usually achieved via expert manual segmentation although automated and semi-automated methods are increasingly providing viable alternatives. Quantitative imaging of cartilage has helped to provide new insights into OA pathogenesis and the influence of various risk factors on disease progression(58–61). Due to the high spatial resolution required for quantitative analysis of cartilage, dedicated MR sequences are required. Typically, these are 3-dimensional gradient echo-based sequences with fat suppression and either fluid suppression (e.g. spoiled gradient echo, SPGR) or fluid excitation (e.g. dual-echo in steady-state, DESS) to optimise contrast resolution between subchondral bone, articular cartilage and synovial fluid. Although these are not part of most clinical MR knee protocols, they are product sequences (i.e. they do not require a specialised research agreement for access) on all major MR vendors and so can be implemented widely. Another morphological quantitative approach is the measurement of subchondral bone area and shape. OA has long been associated with changes in bone shape on plain radiographs(62). However, the assessment of bone area and shape using 3D MR data has provided rich new information on the nature of these changes(63). Head-to-head comparison with quantitative cartilage morphology has demonstrated superior responsiveness for subchondral bone area(64). This can be quantified using the same MR sequences used for cartilage morphology and so is similarly widely applicable. Both quantitative assessment of cartilage thickness and subchondral bone area have been implemented successfully in clinical trials(65,66).

Physiological quantitative approaches are less widespread than their morphological counterparts. This is mostly due to a need for specialised MR pulse sequences (often

only accessible via a research agreement with a vendor and requiring MR physics support) and both technically demanding and time-consuming image analysis pipelines. Their promise lies in an ability to detect and quantify pre-morphological changes and thus the earliest stages of OA (Figure 1.3).

Figure 1.3

Illustration of stages of OA development and progression. Compositional MRI methods offer the potential to detect the very earliest stages of disease before structural changes have become established. Adapted from (67).



From a clinical trials point of view, it also seems logical that physiological changes in response to treatment will occur earlier than morphological changes. The most established physiological imaging methods in OA research are cartilage compositional MR imaging techniques. Such techniques rely, directly or indirectly, on the fact that alterations in cartilage composition will lead to changes in MR relaxation mechanisms. These methods are the focus of chapter 2 of this thesis. Cartilage compositional MR techniques have been available for several decades but their use remains limited largely to the research setting. In common with other advanced imaging methods in OA, this likely reflects the fact that the results do not (at present) influence patient management(68). Other physiological imaging approaches include dynamic contrast-enhanced (DCE) MR imaging. DCE MR imaging quantifies tissue microvascular structure and function via pharmacokinetic modelling of dynamically acquired MR data following the administration of contrast agent. This was originally applied to assess tumour vascularity in oncology studies(69). However, the technique is

potentially applicable to any disease in which tissue vascularity is altered. In OA, it has been applied to quantify the vascularity of both subchondral bone and synovium(70,71). Cross-sectional associations have been demonstrated between DCE parameters of both subchondral bone and synovium and osteoarthritis severity. DCE MR of the synovium has also demonstrated the ability to assess response to therapy with intra-articular steroid(72). DCE MR imaging uses widely available MR pulse sequences, although the post-processing is complex(73).

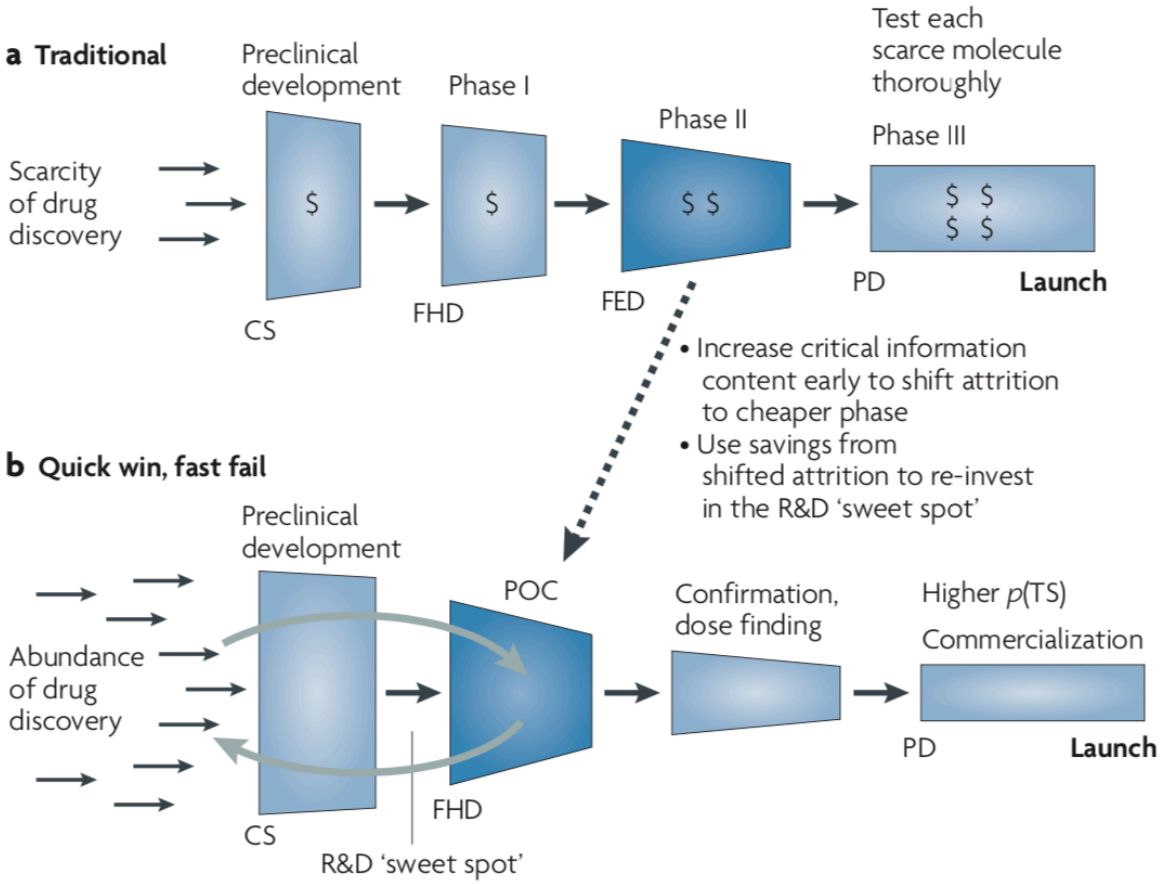
1.3 EXPERIMENTAL MEDICINE

1.3.1 Concept

Experimental medicine is defined by the Medical Research Council as "investigation undertaken in humans, relating where appropriate to model systems, to identify mechanisms of pathophysiology or disease, or to demonstrate proof-of-concept evidence of the validity and importance of new discoveries or treatments"(74). From a treatment development point of view, the purpose of experimental medicine studies is to move the point at which proof-of-concept (POC) or proof-of-mechanism (POM) can be demonstrated for a new drug (or other treatment) to earlier in the development process (Figure 1.4).

Figure 1.4

Comparison of traditional model of drug development (a) with experimental medicine paradigm (b). In (b), the aim is to improve confidence in novel treatments before the more expensive and time-consuming later phases of drug development. **Abbreviations:** CS – candidate selection, FHD – first human dose, FED – first efficacy dose, PD – product decision, POC – proof-of-concept, p(TS) – probability of success. Reproduced from (75).



The main impetus behind the development of the experimental medicine concept has been increasingly high rates of attrition late in the drug development pathway with associated cost implications for pharmaceutical companies(75). The aim of the experimental medicine model is to allow candidate treatments which demonstrate early POC to be backed confidently with the required resource for subsequent stages in drug development. On the other hand, treatments which fail to demonstrate POC can be terminated early, reducing waste(76). Experimental medicine activities rely on the availability of good biomarkers and surrogate outcome measures. The definitions of the terms 'biomarker' and 'surrogate outcome' overlap to some extent. A biomarker is a measured objective characteristic which is an indicator of normal biological processes, pathogenic processes, or a response to a therapeutic intervention. A surrogate outcome measure is defined as a marker that is thought to predict clinical benefit, but is not itself a measure of clinical benefit(77). Biomarkers can be divided into different

categories, for example diagnostic biomarkers aim to discriminate between the disease state of interest and health whereas predictive biomarkers aim to predict the likelihood that a given individual will develop the disease state of interest following an exposure. Biomarkers can include laboratory measurements, physical signs and imaging features.

1.3.2 Imaging biomarkers

Imaging biomarkers are biomarkers measurements derived from medical images. Quantitative imaging biomarkers (QIBs) are a subset of imaging biomarkers where the biomarker is a measurement on a continuous, interval or ratio scale(78). Categorical and qualitative imaging biomarkers are thus excluded from this definition.

QIBs offer the potential to improve precision and reduce subjectivity of image analysis. Compared to qualitative imaging biomarkers, QIBs offer a larger dynamic range and improved sensitivity to change. The implementation of appropriate validated QIBs may allow the early demonstration of POC which is the *raison d'être* of experimental medicine. However, there are challenges specific to the implementation of QIBs compared with laboratory biomarkers. These include the fact that clinical imaging platforms are designed to provide diagnostic quality images for qualitative radiological interpretation, not to provide robust quantitative data. Moreover, QIBs often aim to reflect a complex biological process rather than a single molecule or analyte making straightforward biological validation impossible.

Despite these drawbacks, QIBs are playing an increasingly important role in the drug development process. Collaborative efforts such as the FDA Critical Path Initiative and the creation of the FDA-NIH (National Institutes for Health) Biomarker Working Group have established standardised pathways for biomarkers from initial development and validation through to eventual regulatory acceptance(79,80). Several imaging biomarker-specific initiatives and guidelines have also been developed. These include an 'imaging biomarker roadmap' consensus document which, although written with oncological imaging biomarkers in mind, has broad applicability across disease areas(81). A Quantitative Imaging Biomarkers Alliance has also been created, aiming to standardise QIB terminology and help ensure that correct and consistent statistical methods are applied in QIB studies(82,83).

1.3.3 Relevance to osteoarthritis

Several putative disease modifying OA drugs (DMOADs) have failed in the late stages of drug development. As well as the associated resource implications, this has also discouraged further pharmaceutical industry investment in this area. The EM paradigm therefore has a significant contribution to make to future DMOAD development by providing early evidence of efficacy, thus ensuring that future treatments progress to late stage clinical trials with a greater likelihood of success(84,85).

As described above, QIBs are likely to play an important role in EM studies of OA. There are particular reasons to prefer imaging biomarkers over blood/serum biomarkers in the setting of OA due to the lack of joint specificity of the latter. Moreover, there is increasing recognition that early changes in structure may be predictive of longer-term clinical outcomes. In EM studies, such clinical outcomes (e.g. patient reported outcome scores) are difficult to assess due to the large expected magnitude of placebo effect(86). This has been recognised recently by the FDA with the prospect of obtaining accelerated approval for potential DMOADs based on an imaging-derived surrogate outcome, with subsequent post-approval studies to confirm benefit on clinical outcomes(87). As well as demonstrating early POC, QIBs may also contribute to decision making for later phase studies by informing outcome measurement and recruitment strategy (e.g. recruiting participants based on a QIB threshold).

1.4 THESIS AIMS

More work is required for the promise of QIBs in EM studies to be realised. While several QIBs (particularly cartilage thickness) have been deployed successfully in large observational studies and large clinical trials, there is a relative paucity of evidence on their utility in the EM setting. This is important as the desirable performance characteristics are different in these two settings. For example, in later stage trials the multi-centre feasibility of QIBs is much more of a consideration than in EM studies which may be single-centre or, if multi-centre, involve only a small number of sites.

At present, this lack of evidence hinders the widespread use of QIBs for EM studies. Thus, the purpose of this thesis is to improve the understanding of MR-based QIBs of knee OA in general, and in particular provide new data on their performance characteristics for EM studies.

In chapter 2, I systematically review the literature on the assessment of cartilage composition using MR imaging, and provide a meta-analysis of reliability and discriminative validity of the various techniques described, pooling data from 1,989 persons.

Chapter 3 describes the development and validation of 3D Cartilage Surface Mapping (3D-CaSM), a novel semi-automatic analysis pipeline for articular cartilage which overcomes some of the problems with existing methods described in chapter 2. This validation is performed in five cadaveric knees.

Moving from articular cartilage to subchondral bone, chapter 4 describes the application of texture analysis to 122 persons in the Osteoarthritis Initiative, a prospective multicentre observational study. I assess the association of subchondral bone texture on MR imaging with the radiographic worsening of OA over a 3-year period, as well as the ability of subchondral bone texture to predict this worsening.

Chapters 5 to 8 describe the Advanced Magnetic Imaging of Osteoarthritis (AMROA) study, a prospective observational clinical study involving 20 persons. Fourteen participants with mild-to-moderate knee osteoarthritis and six age-matched healthy controls underwent multiple knee MR examinations over a 1-year period. The aim of this study was to assess the performance characteristics of a number of QIBs which have the potential to contribute to decision making in EM studies. Reflecting the whole-organ concept of knee OA, QIBs of synovium, cartilage and subchondral bone were included. The test-retest repeatability, ability to discriminate between OA participants and healthy controls, and responsiveness to change over 6-months and 1-year were assessed. The results enable head-to-head comparison of the QIBs assessed and will be informative for their inclusion in future longitudinal and interventional studies. The thesis closes with a summary of the work performed with reflection on knowledge gained and directions for future work.

CHAPTER 2:

SYSTEMATIC REVIEW AND META-ANALYSIS OF THE RELIABILITY AND DISCRIMINATIVE VALIDITY OF CARTILAGE COMPOSITIONAL MRI IN KNEE OSTEOARTHRITIS

MacKay, J. W. et al. Systematic review and meta-analysis of the reliability and discriminative validity of cartilage compositional MRI in knee osteoarthritis. Osteoarthr Cartil 26, 1140–1152 (2018).

Cartilage compositional MR techniques aim to quantify changes in the collagen and proteoglycan content of articular cartilage, offering the potential to detect very early degeneration. Here I systematically review the literature on these techniques and perform meta-analysis to assess their reliability and ability to discriminate between subjects with OA and controls. I show that the techniques are generally reliable and – in some cases – able to discriminate between subjects with OA and controls. However, there are several limitations to the current literature leading to increased risk of bias, primarily the use of poorly matched control groups. Moreover, the magnitudes of the differences between OA subjects and controls using existing approaches are small.

2.1 INTRODUCTION

2.1.1 Background

Breakdown of articular cartilage is an important feature of knee osteoarthritis (OA). The earliest changes in articular cartilage are alterations in the biochemical composition of the extracellular matrix (ECM), a network of collagen fibrils and glycoproteins. These compositional changes may predispose to the development of focal defects, which in turn may lead to more diffuse cartilage loss associated with established OA.

Cartilage compositional MR imaging techniques such as T1rho relaxometry, T2 relaxometry and delayed gadolinium enhanced MRI of cartilage (dGEMRIC) are sensitive to changes in cartilage ECM composition, and provide a way to detect degeneration before gross morphological changes become apparent. This contrasts with conventional clinical MR imaging which can detect focal defects and diffuse cartilage loss but is limited in its ability to detect earlier changes in cartilage composition. Compositional MR imaging techniques may therefore allow identification of individuals suitable for intervention at an earlier stage, before

irreversible changes occur. They also have the potential to assess response to treatments designed to repair or regenerate cartilage or slow degradation(88).

2.1.2 Cartilage compositional MRI techniques

A detailed review of the relaxation mechanisms underpinning the different compositional methods is beyond the scope of this thesis. However, an outline of the main compositional methods currently in use is provided below.

T1rho relaxation refers to longitudinal (spin-lattice) relaxation under the influence of a radiofrequency field. It is sensitive to interactions between restricted motion (i.e. bound) water molecules and adjacent macromolecules and is therefore sensitive to the macromolecular content of tissues(89). In articular cartilage, T1rho has been shown to be sensitive to proteoglycan content, although the strength of reported correlation between T1rho relaxation times and proteoglycan content varies between studies(90,91).

T2 relaxation refers to transverse (spin-spin) relaxation. It is sensitive to the ability of water molecules to move and exchange energy freely. For example, fluid-filled structures will tend to have high T2 values due to the presence of freely mobile water molecules. In articular cartilage, T2 has shown to be a non-specific marker of cartilage health, sensitive to changes in water content, collagen content and collagen orientation(92–94). As with T1rho, the strength of correlation between T2 relaxation times and these measures varies between studies.

dGEMRIC refers to the assessment of spin-lattice (T1) relaxation following the administration of gadolinium-based contrast agent (gdCA). dGEMRIC relies on the use of a negatively charged gdCA. Healthy articular cartilage has a bulk negative charge primarily due to the presence of negatively charged glycosaminoglycan (GAG) side chains of proteoglycans. In diseased cartilage, there is a decrease in GAG concentration, with a corresponding decrease in the bulk negative charge. Therefore, more negatively charged gdCA will be allowed to enter diseased cartilage than healthy cartilage, reflected in shorter T1 relaxation times in areas of GAG-depleted cartilage(95). For dGEMRIC to work, gdCA must be able to enter the joint and equilibrate between the synovial and articular cartilage compartments. dGEMRIC achieves this via the intravenous administration of gdCA with eventual 'leakage' of

gdCA into the joint, a process typically encouraged by exercising the joint to be imaged. In order for gdCA to equilibrate between the synovial fluid and articular cartilage, a waiting period (typically 90 minutes following intravenous gdCA) is needed between the administration of gdCA and acquisition of the dGEMRIC images(96). dGEMRIC has been shown to be relatively specific to GAG content(97).

Sodium imaging directly images the MR-active sodium nucleus (^{23}Na), requiring specialist hardware and software for non-proton imaging. Healthy articular cartilage contains cations, principally sodium ions, which are electrostatically attracted by the aforementioned negatively charged GAG side chains. Therefore, the concentration of sodium within articular cartilage is proportional to its GAG content. Sodium imaging has demonstrated high sensitivity and specificity for the detection of GAG depletion in diseased cartilage, although there are limitations including marked reduction in signal-to-noise ratio and longer acquisition times when compared to proton imaging(98).

2.1.3 Rationale for current work

Previous systematic reviews have assessed the reliability and discriminative validity of radiographic and conventional clinical MR imaging assessment of knee osteoarthritis(49,51,53). However, there has been no systematic review which has evaluated the reliability or discriminative validity of cartilage compositional techniques. This was identified as a gap in the literature in recent Osteoarthritis Research Society International (OARSI) guidelines for the use of imaging in the setting of OA clinical trials(44).

As cartilage compositional MR imaging techniques grow in popularity it is important to understand how reliable the techniques are and how well they can distinguish cartilage in individuals with OA compared to cartilage in healthy controls. Accordingly, the aim of this study was to assess the reliability and discriminative validity of cartilage compositional MR imaging in knee osteoarthritis.

For the purposes of this review, I use the term “reliability” to encompass both repeatability (measurement precision with conditions remaining unchanged between repeat measurements e.g. same observer, same MR platform) and reproducibility (measurement precision with conditions changing between repeat measurements, e.g. change in observer, change in MR platform) as defined by the Radiological Society of

North America Quantitative Imaging Biomarkers Alliance (RSNA-QIBA) Metrology Working Group(82).

2.2 METHODS

This systematic review and meta-analysis was undertaken in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations(99).

2.2.1 Protocol & Registration

The study review protocol was prospectively registered on PROSPERO, the international prospective register of systematic reviews (available at www.crd.york.ac.uk/prospero/display_record.asp?ID=CRD42016045250).

2.2.2 Eligibility criteria

In-vivo studies in human subjects involving at least one cartilage compositional MR imaging technique at the knee were considered. The list of compositional techniques considered included, but was not limited to, T1rho relaxometry, T2 relaxometry, T2* relaxometry, dGEMRIC, sodium imaging, glycosaminoglycan chemical exchange saturation transfer imaging (gagCEST), diffusion weighted imaging (DWI) and diffusion tensor imaging (DTI). For a study to be included, it needed to provide reliability data on the technique used (either in subjects with OA or healthy controls or both) or provide measurements comparing subjects with OA to a control group (i.e. discrimination data) or provide both reliability and discrimination data. Only full-text papers reporting original data were included. Conference abstracts, review papers, letters to the editor and opinion pieces were excluded. Included studies were limited to those published in English. Studies using animal models or human tissues *ex-vivo* were excluded.

2.2.3 Information sources

MEDLINE (1946 – February 2017) and EMBASE (1974 – February 2017) were searched via OVID. The databases OpenGrey, Clinicaltrials.gov and the World Health Organization (WHO) International Clinical Trials Registry were also searched for additional studies. The search strategy for the OVID search is presented in Appendix 2.1. The reference lists of full-text manuscripts obtained, personal databases and the contents Tables of key journals were scrutinised for any omitted studies.

2.2.4 Study selection

Two researchers performed initial screening to identify potentially eligible studies per inclusion and exclusion criteria. The full-texts of all potentially eligible papers were then evaluated to enable a final decision on inclusion. Discrepancies were resolved by discussion between the reviewers.

2.2.5 Data extraction and list of items

Data extraction was performed by a single researcher using a piloted electronic data collection form and subsequently verified by a second researcher, with disagreements resolved by discussion. Where a study was considered potentially eligible but data were not presented in an extractable format (e.g. presented in a figure without raw values), the corresponding study author was contacted by email to attempt to obtain the relevant data.

Data extracted for all studies included the following: year of publication, number of participants, age and sex of participants, study design, definition of OA used by the study (if applicable), details of the MR acquisition protocol, MR field strength, experience and training of image analysts, blinding of image analysts to additional clinical information (e.g. OA/control status) and type of regional or subregional analysis performed.

Study participants with OA were divided into those with mild OA (Kellgren-Lawrence (KL) grades 1-2), severe OA (KL grades 3-4) or OA not otherwise specified (NOS) when the study did not provide the information required to stratify(27).

2.2.6 Risk of bias in primary studies

The risk of bias for studies of reliability was performed using a modification of the Quality Appraisal of Diagnostic Reliability (QAREL) tool relevant to this analysis(100). Assessment of risk of bias in studies of discriminative validity was performed using a modification of the revised Quality Assessment of Diagnostic Accuracy (QUADAS-2) tool(101). Full details of the modifications made to QAREL and QUADAS-2 tools are presented in Appendix 2.4.

2.2.7 Data analysis

The primary endpoint for the reliability assessment was a narrative summary of the reliability statistics for intra and inter-observer and test-retest reliability.

A meta-analysis was not appropriate due to the heterogeneity in methods for calculating reliability metrics from the included studies. For example, the coefficient of variation (CV) may be presented as a single value, or the root-mean-square average of several values (RMSCV). There are numerous approaches to computing the intraclass correlation coefficient (ICC) which prevent pooling, and directly comparing ICC across different populations could be misleading(102,103). For the purposes of this review I use an interpretation of ICC values based on Landis and Koch(104).

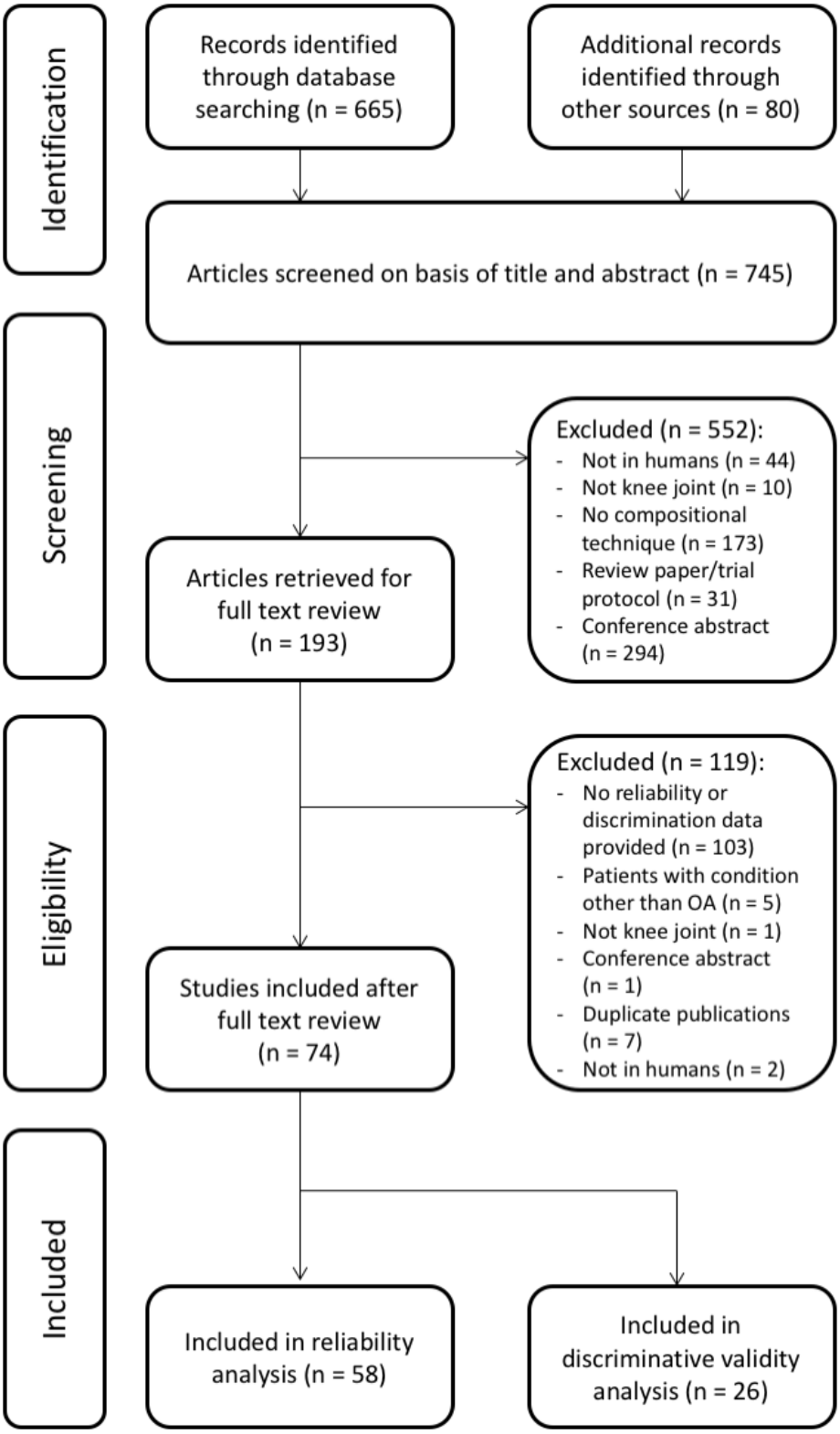
The primary endpoints for the discriminative validity part of this review were estimates of standardised mean difference (average difference between groups divided by the pooled standard deviation of the two groups, analogous to effect size) between subjects with OA and normal controls for each compositional technique studied. Secondary endpoints were estimates of standardised mean difference between subjects with OA and normal controls for each compositional technique studied, limited to studies where control group participant age had been matched to the OA group. The standardised mean difference was used to allow comparison across different compositional techniques with values which vary considerably in magnitude. Where there were less than two studies available for any given comparison, a narrative analysis was conducted.

The appropriateness for meta-analysis was evaluated by assessing the data extraction table for study heterogeneity in cohort characteristics, imaging technique, analysis technique and study processes. Where study heterogeneity was evident for one or more of these factors, a narrative analysis was undertaken. When this did not occur, a meta-analysis was undertaken. In each analysis, statistical heterogeneity was calculated through the I^2 statistic. Fixed effects models were used to pool outcome measures with low heterogeneity ($I^2 \leq 10\%$), whereas random effects models were used to pool outcome measures with high heterogeneity ($I^2 > 10\%$). A strict I^2 threshold was used, wishing to minimise the risk of any 'unknown' heterogeneity from influencing the interpretation of the analyses, particularly as an emphasis had been placed on excluding 'known' heterogeneity in the assessment of study characteristics.

All meta-analyses were performed using Review Manager version 5.3 (The Cochrane Collaboration)(105).

Figure 2.1

PRISMA flow chart demonstrating process for selection of included studies.



2.3 RESULTS

2.3.1 Study selection

Database searching identified 665 citations, with an additional 79 citations identified through other sources (personal databases, reference lists of included studies, contents Tables of key journals). The full-text version of 192 articles was retrieved for detailed review. Forty-eight articles were included in the reliability assessment, 16 articles were included in the discrimination validity assessment and 10 articles were included in both evaluations (Figure 2.1)(106–179).

2.3.2 Reliability study characteristics

Characteristics of included studies are reported in Appendix 2.2. Data from 1,473 subjects were included in the reliability analysis. The most commonly used compositional technique was T2 relaxometry, featuring in 36 of 58 (62%) studies. The number of participants in each study ranged from five to 289 (median 20). The mean (standard deviation, SD) age of participants was 46.2 (14.5) years. Fifty-three percent of included subjects were female.

Twenty-two studies reported data on intra-observer reliability, 25 studies reported data on inter-observer reliability and 33 studies reported data on test-retest reliability. Three studies provided multi-centre test-retest reliability data.

2.3.3 Discriminative validity study characteristics

Characteristics of included studies are reported in Appendix 2.3. Data from 766 subjects were included in the discriminative validity analysis. The most commonly used compositional technique was T2 relaxometry, featuring in 17 of 26 (65%) studies. The number of participants in each study ranged from 15 to 152 (median 33). The mean (SD) age of OA subjects was 58.3 (4.9) years compared to 40.9 (11.7) years in control subjects with females representing 56% of OA subjects compared to 52% of controls. Eight studies included subjects with mild OA, three studies included both subjects with mild OA and subjects with severe OA and 15 studies did not stratify OA severity and were considered as OA (NOS) for these analyses.

2.3.4 Risk of bias in primary studies

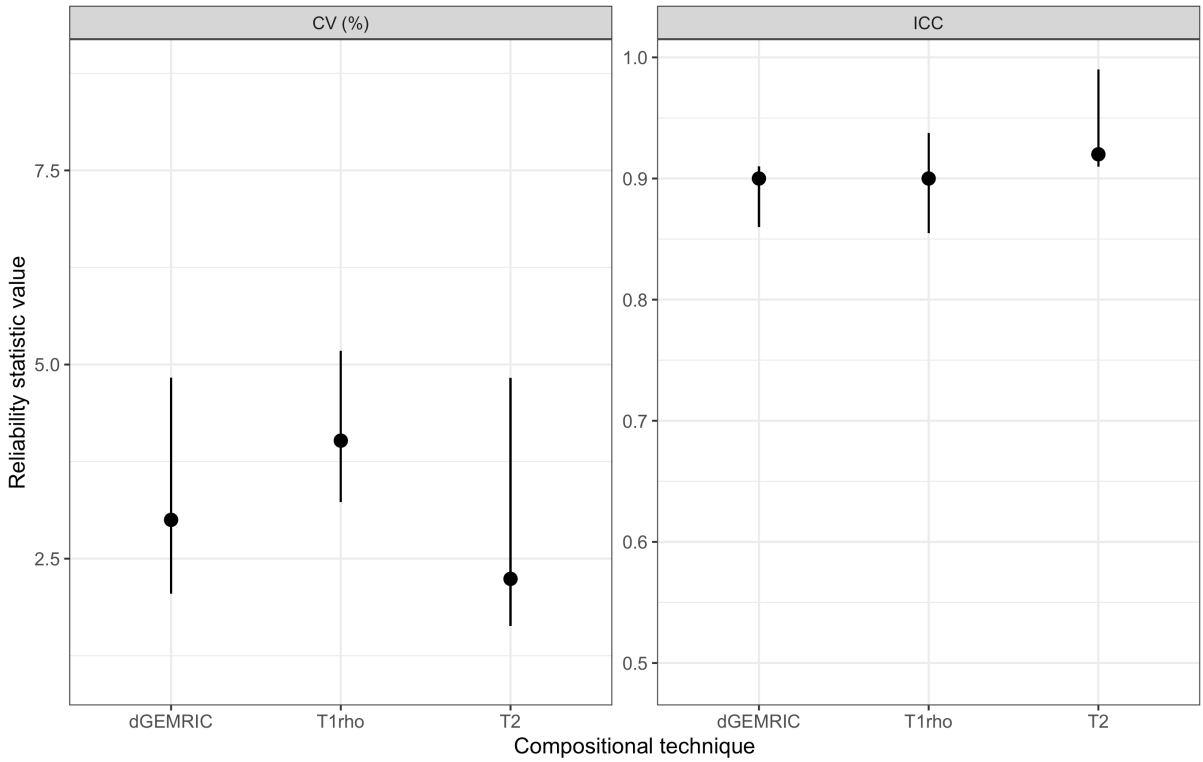
Full results of quality assessments of reliability and discriminative validity studies are presented in Appendix 2.4. Overall, the quality of the evidence was moderate for the reliability assessment and moderate for the discriminative validity assessment. Recurrent weaknesses for the reliability data included the assessment of reliability in only healthy volunteer subjects, lack of information regarding image analyst experience or training, and lack of information regarding image analyst blinding to previous results for studies of intra and inter-observer reliability. Recurrent weaknesses for the discrimination validity data included the use of unmatched control subjects and the potential lack of blinding of image analysts to subject group.

2.3.5 Reliability outcomes

A summary of the results of the reliability analysis are presented in Figure 2.2, with full results in Appendix 2.2.

Figure 2.2

Summary of reliability data (left panel – coefficient of variation (%), right panel – intraclass correlation coefficient) for the three most commonly studied compositional techniques. Central point represents median value, lines represent interquartile ranges. Where a study provided more than one estimate, a mean value was used.



2.3.5.1 T2 relaxometry

Intra-observer ICCs ranged from 0.30 to 0.99 and CVs ranged from 0.8 to 4.7%. Studies featuring multiple subregional analyses tended to report lower ICC values (156). Inter-observer ICCs ranged from 0.17 to 0.99 and CVs ranged from 1.0 to 12.2%. Again, studies which performed analysis on multiple small subregions and analysis of multiple cartilage layers had poorer reproducibility(117,156). CVs for test-retest repeatability where analyses were performed on major compartments (e.g. medial femur, medial tibia etc) were 2.3 to 6.5%, with higher values (up to 22%) again seen where smaller subregional or laminar analyses were performed(117). Three studies examined test-retest repeatability in a multi-centre setting, reporting CVs between 2.3 and 5.3% for major compartments and up to 14% for subregional analyses(107,130,136).

2.3.5.2 T1rho relaxometry

One study provided intra-observer reproducibility data, with a CV of 3.8%(123). All studies of inter-observer reproducibility reported ICC values in the 'excellent' range (> 0.8). Inter-observer CV values ranged from 1.4 to 11.8%. Test-retest repeatability was good-to-excellent (ICCs 0.73 to 0.96, CV 2.3 to 6.1%) when major compartments were analysed but poorer (ICCs as low as 0.2, CV as high as 19%) in two studies where laminar analysis was performed(111,136). Two studies examined test-retest repeatability in a multi-centre setting, reporting CVs of 4.9% for major compartments and up to 18.8% for subregional analyses(130,136).

2.3.5.3 dGEMRIC

Intra and inter-observer reproducibility data were reported by two studies each, with excellent agreement (ICCs > 0.9) and CVs of less than 3%(106,139,158). Test-retest repeatability was excellent (ICCs > 0.85, CV 4.2 to 7.4%) apart from one study comparing different T1 mapping techniques for dGEMRIC which reported ICC values as low as 0 and a CV of 11% for a variable flip angle (VFA) technique(152).

2.3.5.4 Others

Intra- and inter-observer reproducibility data were also reported for sodium imaging, gagCEST, T2* relaxometry, T1 relaxometry (without contrast), magnetisation transfer (MT) imaging and ultrashort TE T2* (UTE-T2*) relaxometry, with excellent reproducibility (ICCs > 0.8) for gagCEST, T2*, MT and UTE-T2*, CVs of 5.1 to 5.9% for T1 mapping and CVs of 8.1 to 11.4% for sodium imaging.

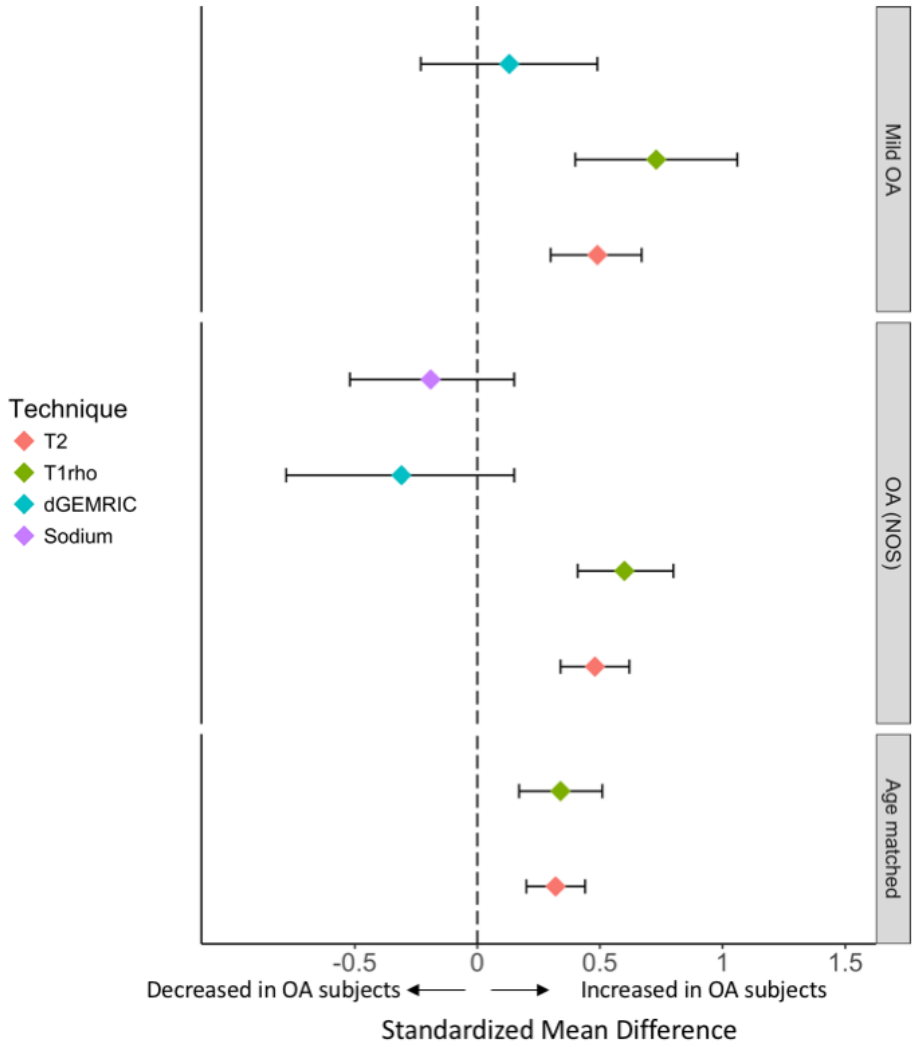
Test-retest repeatability data were reported for the above techniques as well as DTI. Excellent test-retest repeatability ICCs were demonstrated for T2* (0.93) and sodium imaging (0.91)(140,141). Test-retest CVs were generally less than 10% except for sodium imaging which had test-retest CVs between 9.1 and 12.3%(123,134).

2.3.6 Synthesis of results – discriminative validity

Results of the meta-analysis of discriminative validity are presented in Table 2.1 and Figure 2.3. Individual forest plots for each method are presented in Appendix 2.5.

Figure 2.3

Summary forest plot comparing standardised mean differences between subjects with and without OA for each technique where pooling of data was possible.



2.3.6.1 Mild OA

Both T2 and T1rho relaxometry demonstrated significant discrimination between subjects with mild OA and controls ($p < 0.001$), with a greater standardised mean difference (SMD) for T1rho (0.73, 95% CI 0.40 to 1.06) than for T2 (0.49, 0.30 to 0.67). dGEMRIC did not show significant discrimination for mild OA with a SMD of 0.13 (95% CI -0.23 to 0.49, $p = 0.47$). Single studies evaluating MT imaging in patellar cartilage and DTI imaging did discriminate significantly between subjects with mild OA and controls in some compartments(147,153).

2.3.6.2 Severe OA

Pooling of discrimination data was possible for T2 relaxometry only, which demonstrated significant discrimination between subjects with severe OA and controls ($p < 0.001$) with, as expected, a greater SMD of 1.24 (0.63 to 1.85) when compared to the mild OA data. Single studies evaluated discrimination validity for T1rho, dGEMRIC and MT, demonstrating significant differences between groups for T1rho and dGEMRIC but not MT.

2.3.6.3 OA (Not Otherwise Specified)

T2 relaxometry and T1rho relaxometry demonstrated significant differences between subjects with OA (NOS) and controls ($p < 0.001$). As for mild OA subjects, T1rho relaxometry had a higher SMD (0.60, 0.41 to 0.80) than T2 relaxometry (0.48, 0.34 to 0.62). dGEMRIC ($p = 0.18$, SMD = -0.31, -0.78 to 0.15) and sodium imaging ($p = 0.17$, SMD = -0.20, -0.50 to 0.09) did not significantly discriminate OA subjects from controls. Single studies of T2* relaxometry and DTI demonstrated significant discrimination in some compartments(140,148).

2.3.6.4 Additional analyses

When analysis was restricted to studies that had control groups matched to OA subjects for age, data for T2 and T1rho relaxometry comparing OA (NOS) subjects with controls could be pooled. Both techniques retained significant discrimination between subjects with OA and controls ($p < 0.001$), but with lower SMD values of 0.32 (0.20 to 0.44) for T2 and 0.34 (0.17 to 0.51) for T1rho.

Table 2.1

Summary of discrimination between subjects with OA and controls for included compositional techniques. Only techniques with data available for pooling (i.e. at least 2 studies per comparison) are included in this Table.

Compositional technique	**Standardised Mean Difference (95% CI)**					
	Mild OA vs controls	*n*	Severe OA vs controls	*n*	OA (NOS) vs controls	*n*
T2	**0.49 (0.30, 0.67)**	8	**1.24 (0.63, 1.85)**	3	**0.48 (0.34, 0.62)**	10
T1rho	**0.73 (0.40, 1.06)**	4	-	-	**0.60 (0.41, 0.80)**	7
dGEMRIC	0.13 (-0.23, 0.49)	2	-	-	-0.31 (-0.78, 0.15)	3
Sodium	-	-	-	-	-0.19 (-0.52, 0.15)	3

n – number of studies pooled for standardised mean difference estimate
Pooled comparisons in **bold** were statistically significant ($p < 0.05$)

2.4 DISCUSSION

This systematic review has shown that cartilage compositional MR imaging techniques perform well across the domains of intra-observer, inter-observer and test-retest reliability. T2 and T1rho relaxometry demonstrated discrimination validity in mild OA and non-specific OA populations.

Reliability values were generally high across all techniques studied with ICC values in the 'excellent' range (> 0.8) and low CVs (< 5% in most cases). Reliability was slightly poorer for sodium imaging. The reliability values here are commensurate with those for established quantitative measures of joint structure such as cartilage volume and quantitative imaging biomarkers in other body systems(51,180,181). This suggests that cartilage compositional MR imaging is suitably reliable for use in the assessment of knee OA, particularly for the techniques of T2 and T1rho relaxometry where there are most data available. Analyses performed on small cartilage subregions or involving laminar analysis (where cartilage is split into 2 or 3 layers from deep to superficial) tended to be less reliable than those assessing larger cartilage regions. This is probably due to the effects of noise and partial volume with adjacent synovial fluid and subchondral bone which are likely to be exacerbated in small regions of interest (ROIs), together with increased scope for subjective positioning differences between observers. This should be borne in mind when such analyses are interpreted. Few studies examined test-retest reliability in a multi-centre setting. To facilitate use of compositional techniques in large-scale clinical trials, more studies in this area are needed.

T1rho and T2 relaxometry demonstrated the ability to discriminate between subjects with mild OA and controls, and subjects with OA (NOS) and controls. T1rho demonstrated larger SMD values in both populations suggesting that it has superior discrimination validity, in keeping with the results of previous *in vivo* and *ex vivo* studies(157,175,182). dGEMRIC and sodium imaging demonstrated smaller and non-significant SMDs between OA subjects and controls. This contrasts with previous work showing better correlation between dGEMRIC and GAG content of articular cartilage than between T1rho values and GAG content(97). Possible reasons for the poorer performance of dGEMRIC in this meta-analysis include variation in imaging protocols

between studies which may have affected results(96), or the fact that dGEMRIC is just a poorer technique *in vivo*.

For all techniques studied, there was significant statistical heterogeneity between different cartilage regions. This concurs with previous work demonstrating significant spatial variation in articular cartilage compositional values at the knee, and potentially suggests that changes in cartilage composition due to OA also show substantial spatial variation(161,167,183).

The principal role for cartilage compositional MR imaging is the detection of adverse changes in cartilage composition prior to morphological damage. Therefore, they are of greatest potential utility in subjects with mild OA and are of questionable value once the disease is more advanced(184). SMD values for T1rho and T2 relaxometry, which showed significant discrimination validity for mild OA population, correspond to relatively small absolute differences between mild OA subjects and controls of 3.5 ms (95% CI 1.9 – 5.2 ms) for T1rho and 1.9 ms (1.2 – 2.7 ms) for T2. Although statistically significant differences have been demonstrated between groups in this meta-analysis, it is questionable whether these differences are of clinical significance, that is of sufficient magnitude to provide useful clinical discrimination when interpreting a single measurement. Moreover, these small differences may be exaggerated because few control groups were matched to OA subjects for important characteristics such as age or sex. Given these small overall differences, novel methods of enhancing the utility of compositional data are being developed. These include novel data-driven analyses such as voxel based relaxometry and cluster analysis which give the ability to detect localised changes without the reliability problems associated with small manual ROIs, and so-called 'functional' cartilage imaging where differences between normal and abnormal cartilage may be exacerbated(146,185–187).

This study has some unavoidable limitations. First, I have considered reliability and discrimination validity of cartilage compositional MR imaging, but not responsiveness to change. This will be an important factor to consider when using cartilage compositional MR imaging as an imaging biomarker of response to treatment, and from a clinical utility point of view it may be that the magnitude of intra-subject change is more important than absolute mean differences between OA and control groups. However, at present, there have not been sufficient studies in this area to

permit pooled analysis. Moreover, the majority of studies which have used cartilage compositional MR imaging in a longitudinal setting do not report sufficient data to allow calculation of standardised response means for pooling. Second, the quality of included studies was variable for both reliability and discriminative validity studies. The reliability values reported across different studies were consistent suggesting that substantial bias affecting the results of this part of the review had not been introduced by these factors. However, discriminative validity results did vary between studies, as indicated by moderate heterogeneity in the meta-analyses. Third, data extraction was performed by a single reviewer only with discussion with a second reviewer in cases of doubt, whereas optimal methodology would involve independent data extraction by two reviewers.

In conclusion, this systematic review and meta-analysis has demonstrated that cartilage compositional MR imaging techniques are reliable and, in the case of T2 and T1rho relaxometry, able to discriminate between subjects with OA and controls.

APPENDIX 2.1

Search strategy for MEDLINE and EMBASE

1	exp osteoarthritis/
2	osteoarth*.tw.
3	or/1-2
4	knee.tw.
5	exp knee/
6	or/4-5
7	exp magnetic resonance imaging/
8	MRI.ti,ab.
9	magnetic resonance.ti,ab.
10	or/7-9
11	map*.tw.
12	relax*.tw.
13	T2.tw.
14	T2*.tw.
15	Tlrho.tw.
16	T1*.tw.
17	sodium.tw.
18	dGEMRIC.tw.
19	gadolinium.tw.
20	gadopentetate.tw.
21	diffusion.tw.
22	DWI.tw.
23	DTI.tw.
24	gagCEST.tw.
25	collagen.tw.
26	proteoglycan.tw.
27	glycosaminoglycan.tw.
28	composition*.tw.
29	or/11-28
30	exp sensitivity/
31	exp specificity/
32	reliab*.tw.
33	valid*.tw.
34	sensitiv*.tw.
35	specific*.tw.
36	discim*.tw.
37	reproduc*.tw.
38	variab*.tw.
39	or/30-38
40	and/3,6,10,29,39

APPENDIX 2.2

Characteristics of included reliability studies

Study	Number of OA subjects			Number of other subjects	Mean age		Gender (% female)		MRI field strength (T)	Compositional techniques used	Type of reliability assessed	Reliability statistic**	Reliability value	Assessors	
	mild	sev	NOS		OA	other	OA	other						Number	Experience
Anandacoomarasamy(106)	-	-	-	20	-	NS	-	NS	3	dGEMRIC	intraO	ICC	0.91	1	NS
Balamoody(107)	-	-	12	-	49	-	25	-	3	T2	TR[#]	RMSCV	2.3 – 6.3	1	NS
Baum(108)	-	-	-	126	-	50	-	50	3	T2	intraO	RMSCV	1.76	1	NS
Blumenkrantz(109)	13	17	-	-	64	-	39	-	1.5	T2	intraO	CV	1.5 – 2	1	NS
Bron(110)	-	-	11	10	52	27	64	40	3	dGEMRIC	TR	ICC	0.85 – 0.9	1	NS
Carballido-Gamio(111)	-	-	-	5	-	29	-	20	3	Tlrho	TR	CV	3.8 – 12.3†	NS	NS
Dardzinski(112)	-	-	5	5	52		70		3	T2	TR	RMSCV	3.3 – 6.5	1	NS
Duryea(113)	-	-	-	10	-	33	-	50	3	T2	intraO TR	CV ICC CV ICC	1.3 0.99 5.8 0.92	1	NS
Guha(114)	-	-	20	20	58	54	50	60	3	DTI	TR	RMSCV	6.5 – 11.6	NS	NS
Gupta(115)	-	-	-	20	-	37	-	35	3	Tlrho	interO	RMSCV ICC	3.9 0.96	2	NS
Hada(116)	50	-	-	19	57	25	54	37	3	T2	intraO interO	ICC ICC	0.91 0.87	2	NS
Hannila 2009(117)	-	-	-	20	-	23	-	50	1.5	T2	interO	RMSCV ICC	1.8 – 14.2† 0 – 0.98†	2	NS
Hannila 2015(118)	-	-	-	9	-	30	-	44	1.5	T2	TR	RMSCV ICC	2.5 – 22.2† 0 – 0.98†	1	NS
Hesper(119)	-	-	-	10	-	29	-	70	3	T2*	intraO	ICC	0.97	1	8 years

Study	Number of OA subjects			Number of other subjects	Mean age		Gender (% female)		MRI field strength (T)	Compositional techniques used	Type of reliability assessed	Reliability statistic**	Reliability value	Assessors	
	mild	sev	NOS		OA	other	OA	other						Number	Experience
Holtzman(120)	-	-	-	26	-	36	-	38	3	Tlrho T2	interO	CV	4.3 4.9	2	At least 25 datasets
Hovis 2011(121)	-	-	-	161	-	50	-	59	3	T2	intraO	RMSE	0.74 – 1.51	1	NS
Hovis 2012(122)	-	-	105	-	65	-	68	-	3	T2	intraO interO	ICC	0.99 0.99	2	NS
Jordan(123)	-	-	-	8	-	28	-	25	3	Tlrho	intraO interO TR	RMSCV	3.8 5.7 4.6 – 6.1	2	NS
										T2	intraO interO TR		4.7 6.7 6.3 – 10.7		
										Sodium	intraO interO TR		8.1 11.4 11.3 – 12.9		
Joseph 2011(124)	-	-	-	145	-	50	-	59	3	T2	intraO	RMSCV	0.9 – 2.1	1	NS
Joseph 2012(125)	-	-	-	289	-	51	-	47	3	T2	intraO	RMSCV	0.8 – 3.2	1	NS
Juras 2016(126)	-	-	-	23	-	33	-	57	3	T2	interO	CV	5.8 – 10.8†	2	10 years/ 15 years
Koli(127)	80	-	-	-	58	-	100	-	1.5	T2	interO	RMSCV	2	NS	NS
X Li 2005(129)	-	-	9	10	52	30	44	40	3	Tlrho	TR	CV	4.8	NS	NS
X Li 2014(128)	-	-	-	6	-	22-35	-	50	3	Tlrho T2	TR	RMSCV	4.2 – 6 5 – 6.3	NS	NS
X Li 2015(130)	-	-	-	18	-	NS	-	NS	3	Tlrho T2	TR[#]	RMSCV	2.3 – 5.1 3.2 – 5.3	NS	NS
Liebl(131)	-	-	-	130	-	59	-	60	3	T2	intraO interO	RMSCV	1.7 1.6	2	NS
Liess(132)	-	-	-	20	-	28	-	20	1.5	T2	TR	CV	1.7	NS	NS
Liu(133)	-	-	14	13	53	28	36	23	3	T2	TR	CV	3.1 - 10	1	4 years

Study	Number of OA subjects			Number of other subjects	Mean age		Gender (% female)		MRI field strength (T)	Compositional techniques used	Type of reliability assessed	Reliability statistic**	Reliability value	Assessors	
	mild	sev	NOS		OA	other	OA	other						Number	Experience
Madelin(134)	-	-	-	6		37	-	50	3, 7	Sodium	TR	RMSCV	9.1 – 12.3	NS	NS
Matsubara(135)	10	-	-	19	57	39	32	0	3	Tlrho	interO	ICC	0.93	2	8 years/ 9 years
Mosher 2011(136)	16	16	-	18	54	27	59	67	3	Tlrho Tlrho T2 T2	TR[#]	ICC RMSCV ICC RMSCV	0.20 – 0.93† 7.23 – 18.83† 0.61 – 0.98† 4.36 – 14†	NS NS	NS Computer tutorial
Mosher 2004(137)	-	-	-	30	-	22-86	-	100	3	T2	intraO	wK	0.77	1	8 years
Multanen 2015(139)	78	-	-	12	59	58	100	100	1.5	T2 dGEMRIC	interO	RMSCV	2 3	2	6 years/ 12 years
Multanen 2009(138)	-	-	-	10	-	32	-	50	1.5	dGEMRIC	TR	ICC RMSCV	0.45 – 0.98† 4.7 – 12. 9†	NS	NS
Newbould 2012(140)	-	-	13	5	64	62	77	80	3	T2*	TR	ICC CV	0.7 – 0.94 3.2 – 7.7	NS	NS
Newbould 2012b(141)	-	-	15	5	64	62	80	80	3	Sodium	TR	ICC CV	0.67 – 0.94 3.6 – 9.9	1	NS
Nishioka 2012(142)	-	20	-	-	77	-	90	-	3	Tlrho T2	interO	CV	< 6 < 6	NS	NS
Nishioka 2015(144)	-	-	78	-	69	-	77	-	3	Tlrho T2	interO	RMSCV	11.8 12.2	2	NS
Nishioka 2013(143)	-	-	-	37	-	23	-	59	3	Tlrho T2	interO	RMSCV	1.4 1.0	2	NS
Pan(145)	-	-	-	95	-	55	-	61	3	T2	intraO	RMSCV	1.1 – 1.2	1	NS
Pedoia(146)	40	-	-	15	54	48	NS	NS	3	Tlrho	TR	CV	2	2	2 years/ 4 years
Raya 2014(147)	5	-	-	10	66	31	NS	30	7	DTI	TR	RMSCV	2.9 – 5.6	NS	NS
Raya 2012(148)	-	-	10	16	61	31	50	44	7	DTI T2	TR	RMSCV	7.3 – 19.3† 5.5 – 6.9†	1	5 years

Study	Number of OA subjects			Number of other subjects	Mean age		Gender (% female)		MRI field strength (T)	Compositional techniques used	Type of reliability assessed	Reliability statistic**	Reliability value	Assessors	
	mild	sev	NOS		OA	other	OA	other						Number	Experience
Schleich(149)	-	-	-	20	-	25	-	40	3	gagCEST	intraO interO	ICC	0.95 0.95	1 2	8 years 5 years/ 8 years
Serebrakian(150)	-	-	-	127	-	55	-	56	3	T2	intraO interO	RMSCV	1.1 3.3	2	NS
Singh(151)	-	-	-	8	-	20-35	-	NS	7	T1rho	TR	ICC CV	0.73 – 0.96 2.3 – 4.3	NS	NS
Siversson(152)	-	-	-	9	-	45	-	56	1.5	dGEMRIC	TR	ICC RMSCV	0 – 0.69 6.0 – 11.6	NS	NS
Sritanyaratana(153)	11	-	-	20	53	32	36	25	3	MT	TR	CV	0.5 – 4.6	1	NS
Stehling 2010(155)	-	-	-	120	-	51	-	50	3	T2	intraO	CV	1.2	1	NS
Stehling 2011(154)	-	-	-	10	-	52	-	50	3	T2	interO	RMSCV	1.2 – 2.8†	2	NS
Surowiec(156)	-	-	-	18	-	18-35	-	NS	3	T2	intraO interO	ICC	0.17 – 0.89 0.3 – 0.96	3	5 years/ 6 years/ 13 years
Takayama(157)	-	16	-	-	73	-	88	-	3	T1rho T2	interO	ICC	0.81 0.92	2	12 years / 7 years
Tiderius(158)	-	-	-	12	-	24	-	0	1.5	dGEMRIC	intraO interO	CV	1.5 – 2.6 1.8	6	2 medical students, 2 ortho surgeons, 2 radiologists
Van Tiel(159)	17	-	-	-	50	-	41	-	3	dGEMRIC	TR	ICC	0.87 – 0.95	1	Medical degree

Study	Number of OA subjects			Number of other subjects	Mean age		Gender (% female)		MRI field strength (T)	Compositional techniques used	Type of reliability assessed	Reliability statistic**	Reliability value	Assessors	
	mild	sev	NOS		OA	other	OA	other						Number	Experience
Welsch(160)	-	-	-	17	-	26	-	24	3, 7	T2	interO	ICC	0.91 – 0.95†	3	2 years/ 10 years/ 25 years
											TR	CV	7.2 – 8.7†		
										T2*	interO	ICC	0.88 – 0.90†		
											TR	CV	6.8 – 7.8†		
										MT	interO	ICC	0.84 – 0.91†		
											TR	CV	9.2 – 10.8†		
Wiener(161)	-	-	-	25	-	31	-	60	1.5	T1	TR	CV	5.1 – 5.9	1	NS
Williams(162)	-	-	-	11	-	28	-	100	3	UTE-T2*	intraO	ICC	0.80 – 0.97†	1	9 years
											TR	RMSCV	6 - 16†		
Zuo(163)	-	-	-	6	-	27-30	-	17	3	T1rho	TR	CV	1.8 – 5.6	NS	NS
										T2			2.9 – 5.7		

Abbreviations: sev - severe, NS – not specified, intraO – intra-observer, interO – inter-observer, TR – test-retest.
**Reliability values are presented as ranges when values were provided separately for different cartilage ROIs. RMSCV and CV values are provided as percentages.
†Laminar (e.g superficial/deep cartilage layers) analysis performed.
[#] Multicentre study

APPENDIX 2.3

Characteristics of included discrimination studies

Study	Number of OA subjects			Number of control subjects	Mean age		Gender (% female)		MRI field strength (T)	Compositional techniques used	Definition of OA
	mild	sev	NOS		OA	control	OA	control			
Dunn(164)	20	28	-	7	63/67*	38	65/50*	43	1.5	T2	Radiographs and symptoms
Eckstein(165)	-	-	77	75	58	55	100	100	3	T2, dGEMRIC	Radiographs only
Hada(116)	50	-	-	19	57	25	54	37	3	T2	ACR criteria
X Li 2005(129)	-	-	9	10	52	30	44	40	3	T1rho	Radiographs and/or symptoms
X Li 2007(168)	-	-	10	16	56	41	30	50	3	T1rho, T2	Radiographs and symptoms
X Li 2009(167)	-	-	10	10	56	41	30	40	3	T1rho, T2	Radiographs and symptoms
W Li 2010(166)	-	-	14	9	62	29	74	64	1.5	dGEMRIC	MRI and/or radiographs and/or symptoms
Liu 2015(133)	-	-	14	13	53	28	36	23	3	T2	Radiographs and symptoms
Madelin(169)	-	-	28	19	64	35	57	42	7	Sodium	ACR criteria
Matsubara(135)	10	-	-	19	57	39	32	0	3	T1rho	Radiographs and symptoms
Mosher(136)	16	16	-	18	51/57*	27	44/75*	67	3	T1rho, T2	Radiographs only
Multanen(139)	78	-	-	12	59	58	100	100	1.5	T2, dGEMRIC	Radiographs and symptoms
Newbould 2012(140)	-	-	20	17	64	61	65	76	3	T2*	ACR criteria

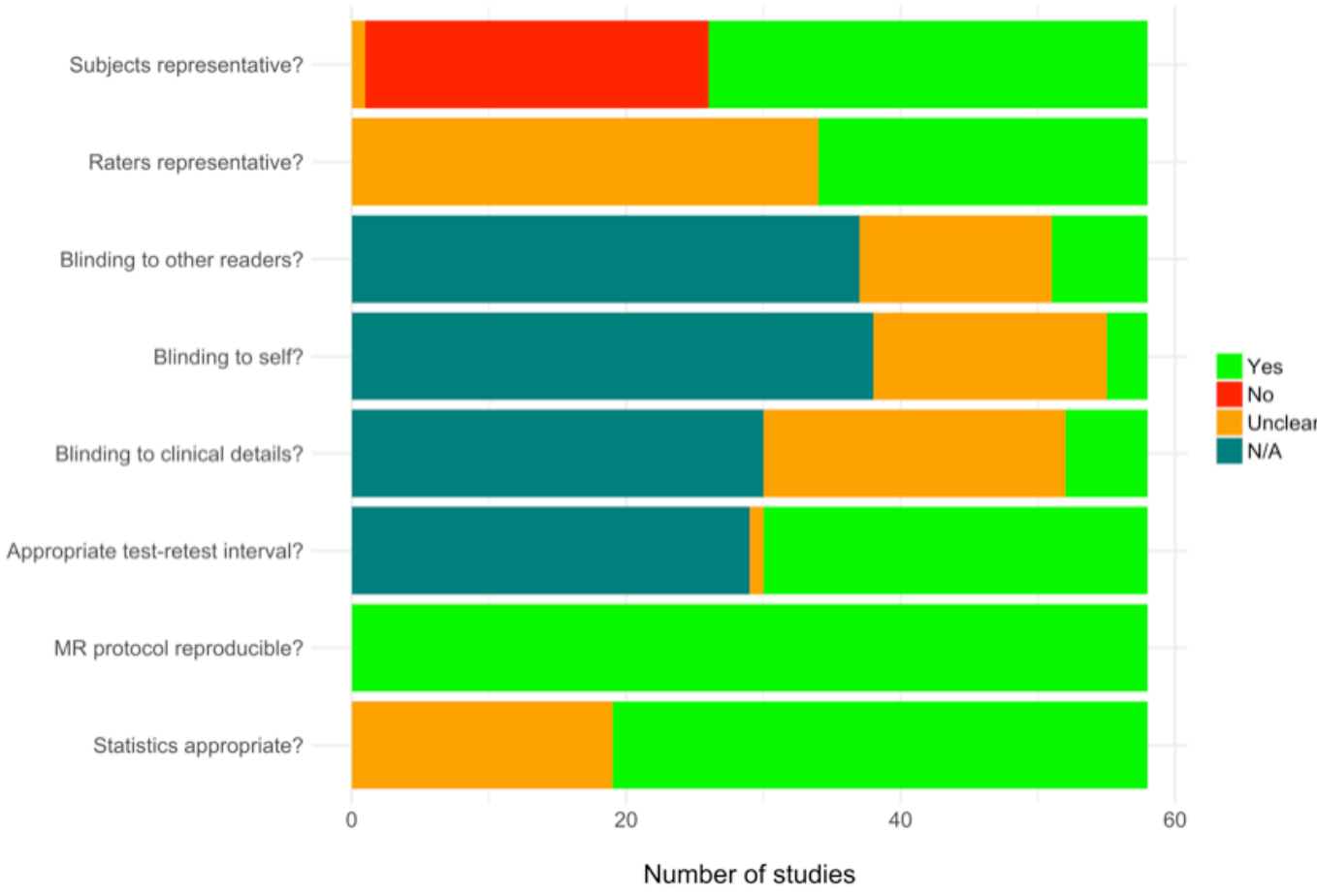
Study	Number of OA subjects			Number of control subjects	Mean age		Gender (% female)		MRI field strength (T)	Compositional techniques used	Definition of OA
	mild	sev	NOS		OA	control	OA	control			
Newbould 2013(178)	-	-	28	23	63	62	NS	NS	3	Sodium	ACR criteria
Owman(170)	-	-	27	7	NS	NS	NS	NS	1.5	dGEMRIC	Radiographs only
Raya 2012(148)	-	-	10	16	61	31	50	44	7	DTI, T2	ACR criteria
Raya 2014(147)	5	-	-	10	66	31	NS	30	7	DTI	ACR criteria
Souza 2013(172)	-	-	44	19	57	39	NS	NS	3	T1rho, T2	Radiographs and symptoms
Souza 2014(171)	-	-	44	93	57	50	61	58	3	T1rho, T2	Radiographs and symptoms
Sritanyaratana(153)	11	-	-	20	53	32	36	25	3	MT	Radiographs and symptoms
Stahl 2007(173)	-	-	8	10	56	58	100	100	3	T2	Radiographs and symptoms
Stahl 2009(174)	17	-	-	20	54	34	53	50	3	T1rho, T2	ACR criteria
Wang(175)	10	-	-	10	65	36	20	40	3	T1rho, T2	Radiographs only
Wirth(176)	32	-	-	89	60	55	56	60	3	T2	Radiographs only
Wyatt(177)	-	-	12	7	59	50	42	71	3, 7	T1rho, T2	Radiographs and symptoms
Yao(179)	20	20	-	11	48/55*	39	65/60*	27	3	T2, dGEMRIC, MT	Radiographs and symptoms

Abbreviations: sev - severe, ACR – American College of Rheumatology, NS – not specified.
*Mean ages and female percentage for mild/severe OA groups provided separately

APPENDIX 2.4

Quality assessment of included studies

1. **Reliability**

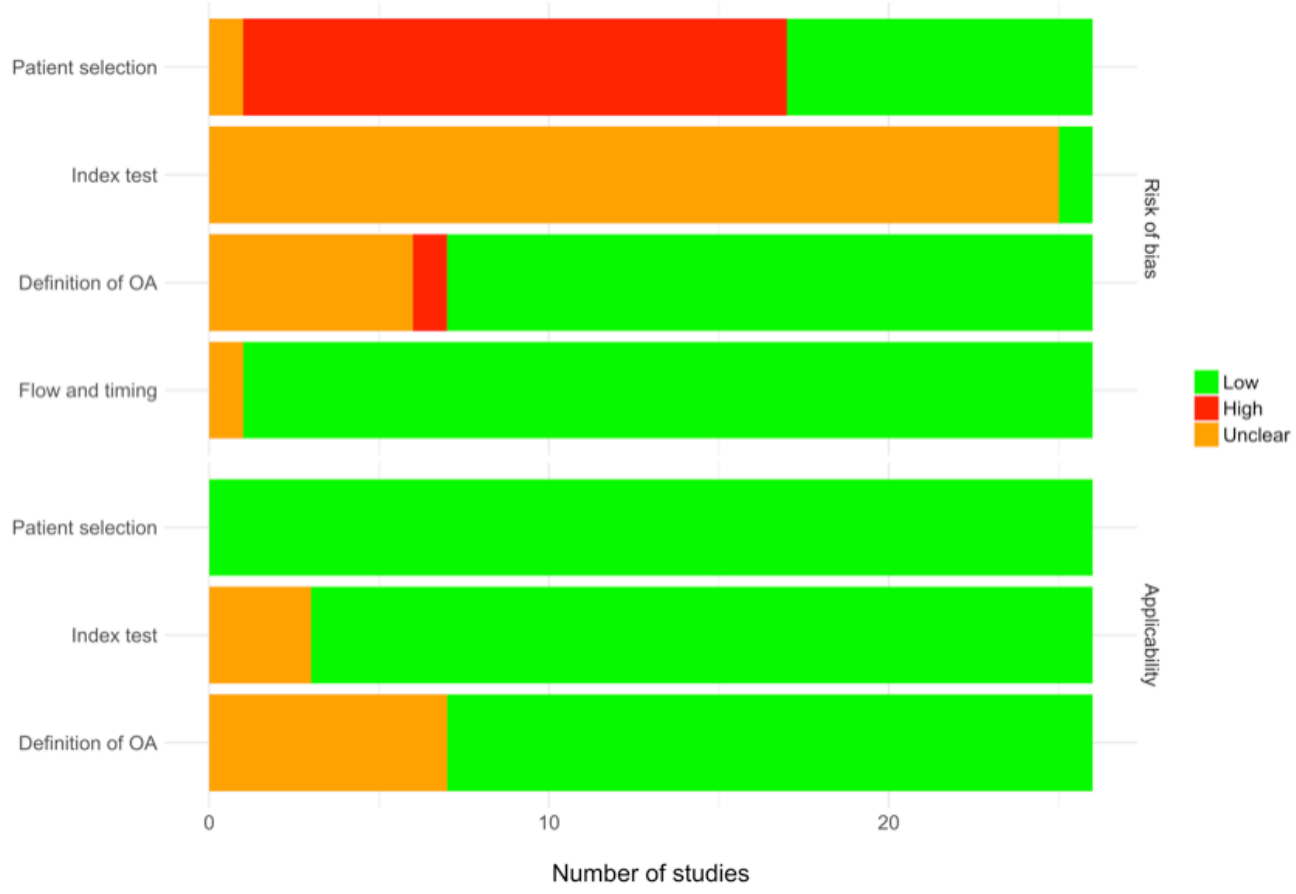


Modified QAREL items as follows:

1. Was the test evaluated in a sample of subjects who were representative of those to whom the authors intended the results to be applied?
2. Was the test interpreted by analysts who were representative of those to whom the authors intended the results to be applied?
3. Were analysts blinded to the findings of other analysts during the study (only inter-observer reliability studies)?
4. Were analysts blinded to their own prior findings of the test under evaluation (only intra-observer reliability studies)?
5. Were analysts blinded to clinical information e.g. age, sex, OA status (only for studies featuring both OA subjects and controls)?

6. Was the time interval between repeated measurements compatible with the stability of the measurement (only test-retest reliability studies)?
7. Was the MR protocol described reproducibly?
8. Were appropriate statistical measures of agreement used?

2. Discriminative validity



Modified QUADAS-2 items as follows:

1. PATIENT SELECTION

Risk of bias

Could the selection of patients have introduced bias?

Signalling question 1: Was an attempt made to match control and OA populations for important baseline characteristics (e.g. age, sex)?

Signalling question 2: Did the study avoid inappropriate exclusions?

Applicability

Are there concerns that the included patients do not match the review question?

2. INDEX TEST

Risk of bias

Could the conduct or interpretation of the index test have introduced bias?

Signalling question 1: Were image analysts blinded to additional clinical cues (e.g. age, sex, OA status)?

Signalling question 2: Is the MR imaging acquisition protocol described in sufficient detail for it to be reproduced?

Signalling question 3: Is the image analysis procedure described in sufficient detail for it to be reproduced?

Applicability

Is there concern that the index test, its conduct or interpretation differ from the review question?

3. DEFINITION OF OA

Risk of bias

Signalling question 1: Was the study definition of OA based on clinical symptoms, imaging findings or both? Were standard (ACR) criteria applied?

Applicability

Is there concern that the definition of OA used differs from the review question?

4. FLOW AND TIMING

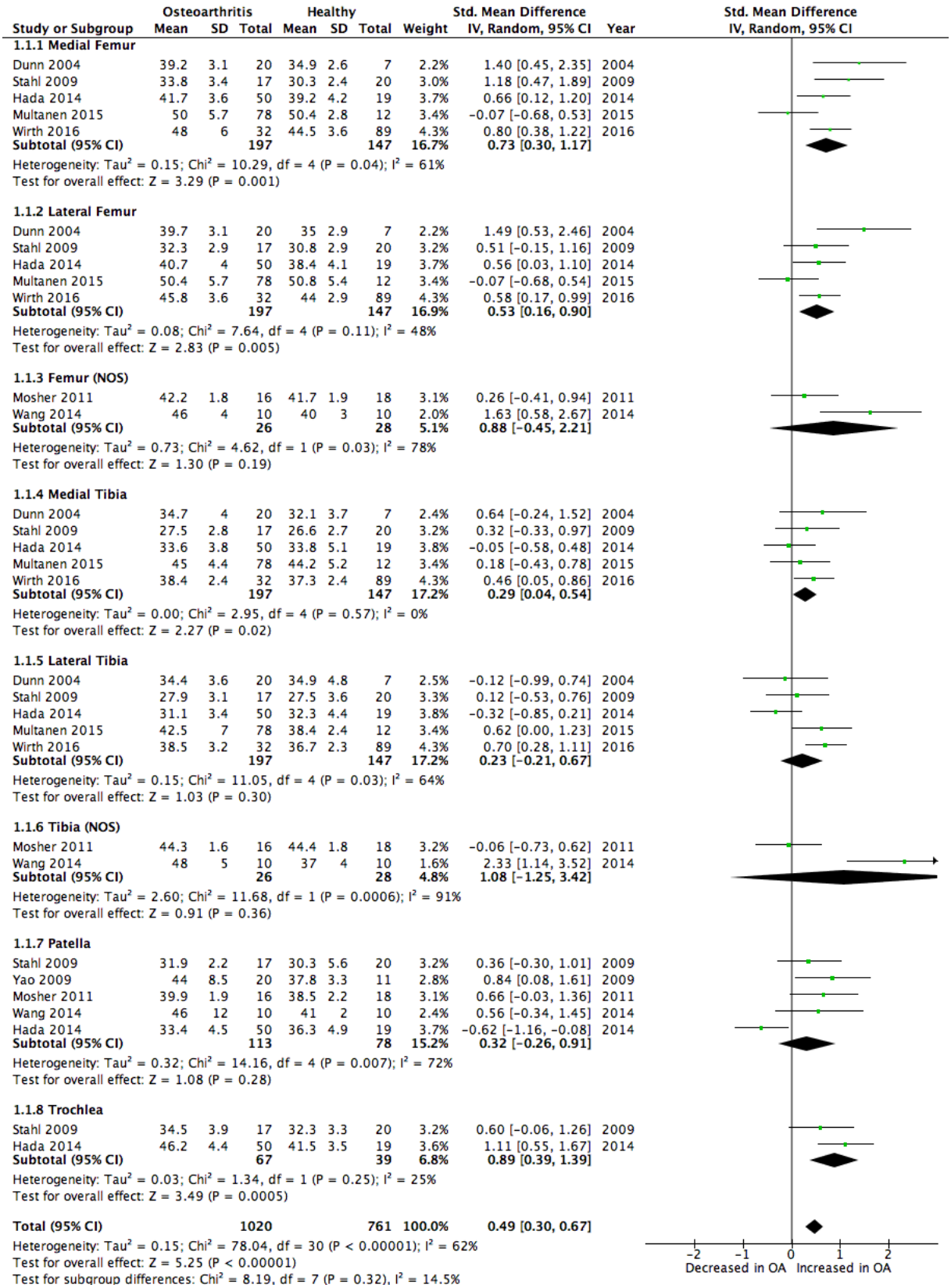
Risk of bias

Signalling question 1: Were all participants included in the analysis?

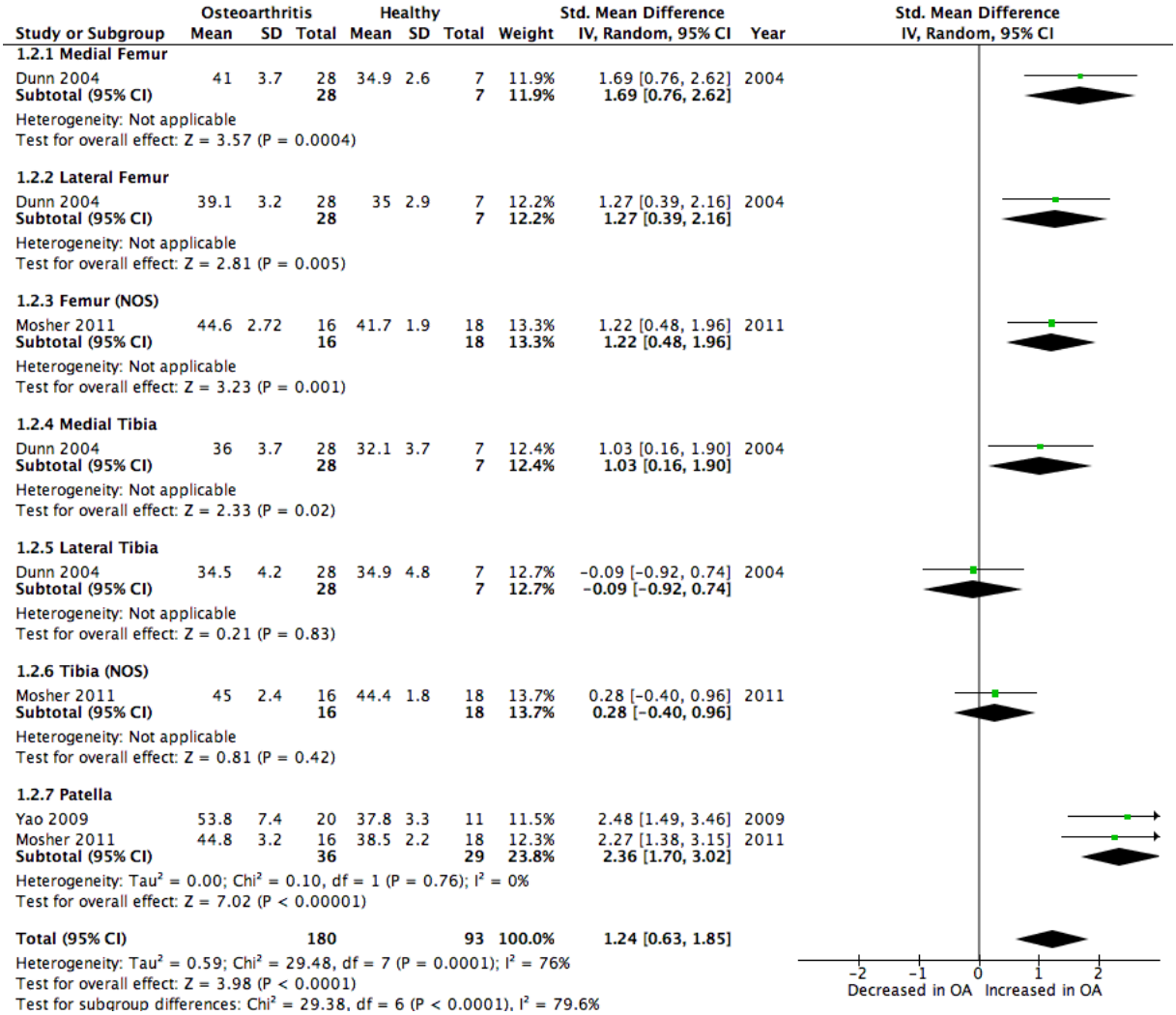
APPENDIX 2.5

Forest plots for discriminative validity assessment

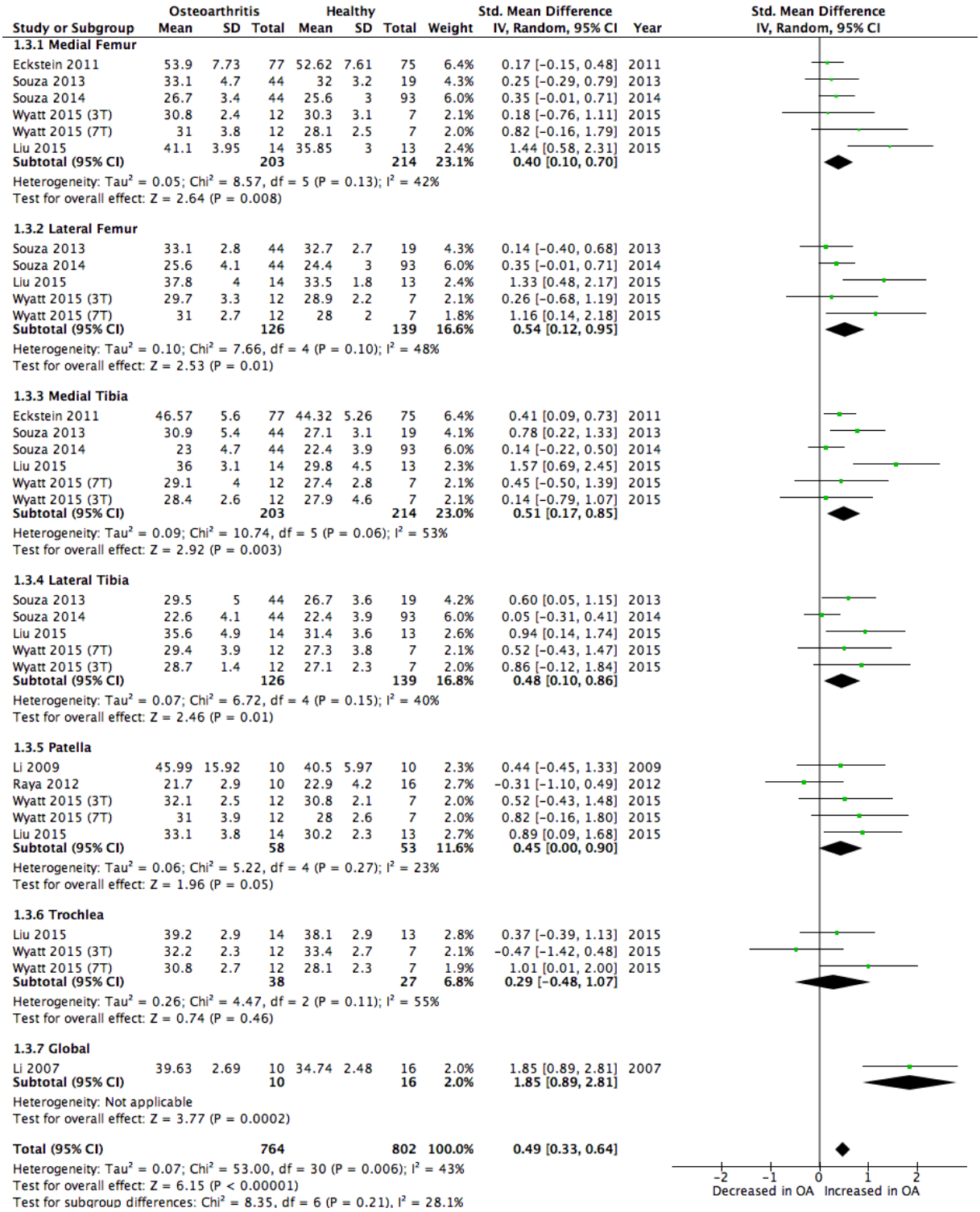
T2: mild OA vs controls



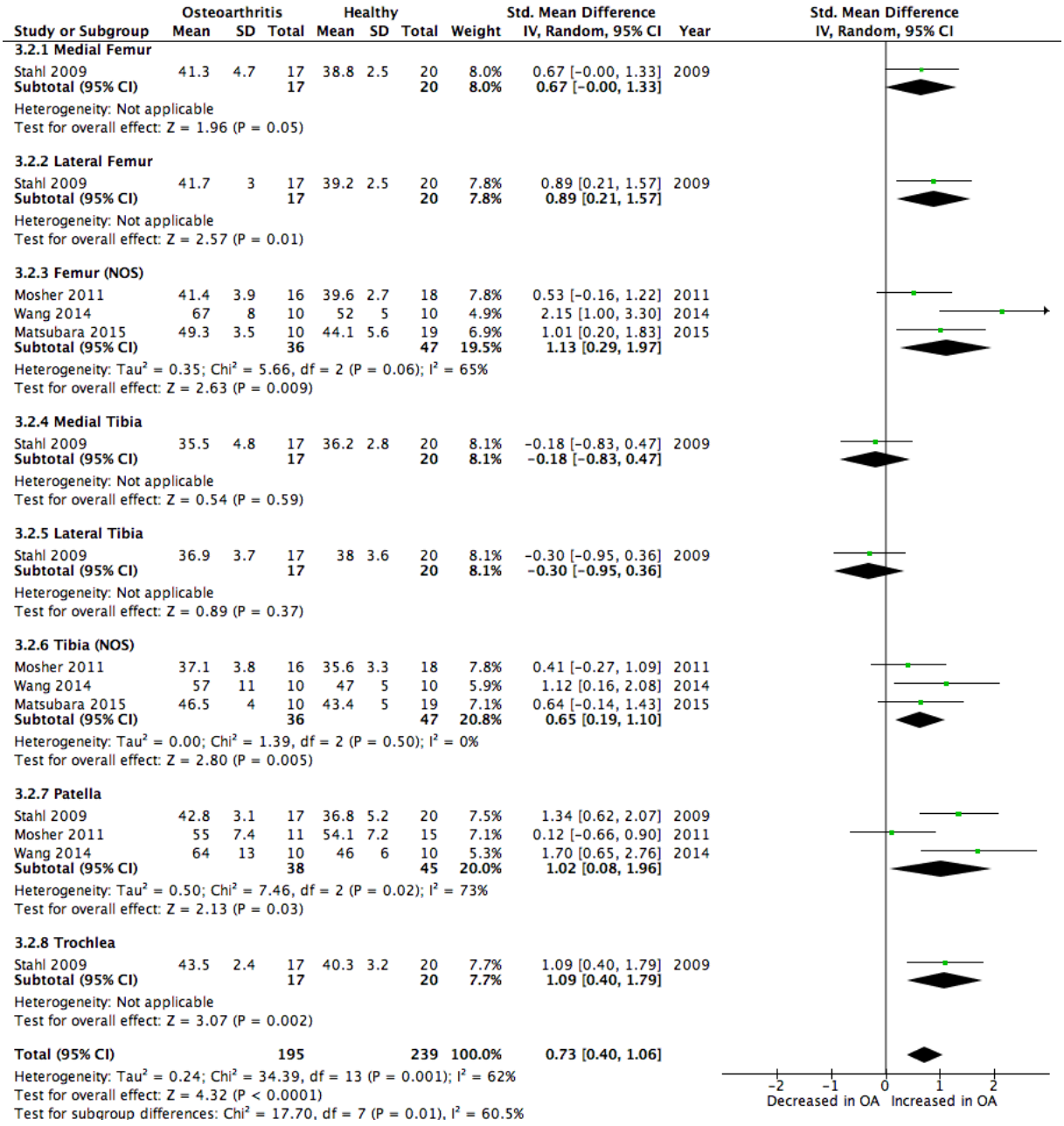
T2: Severe OA vs controls



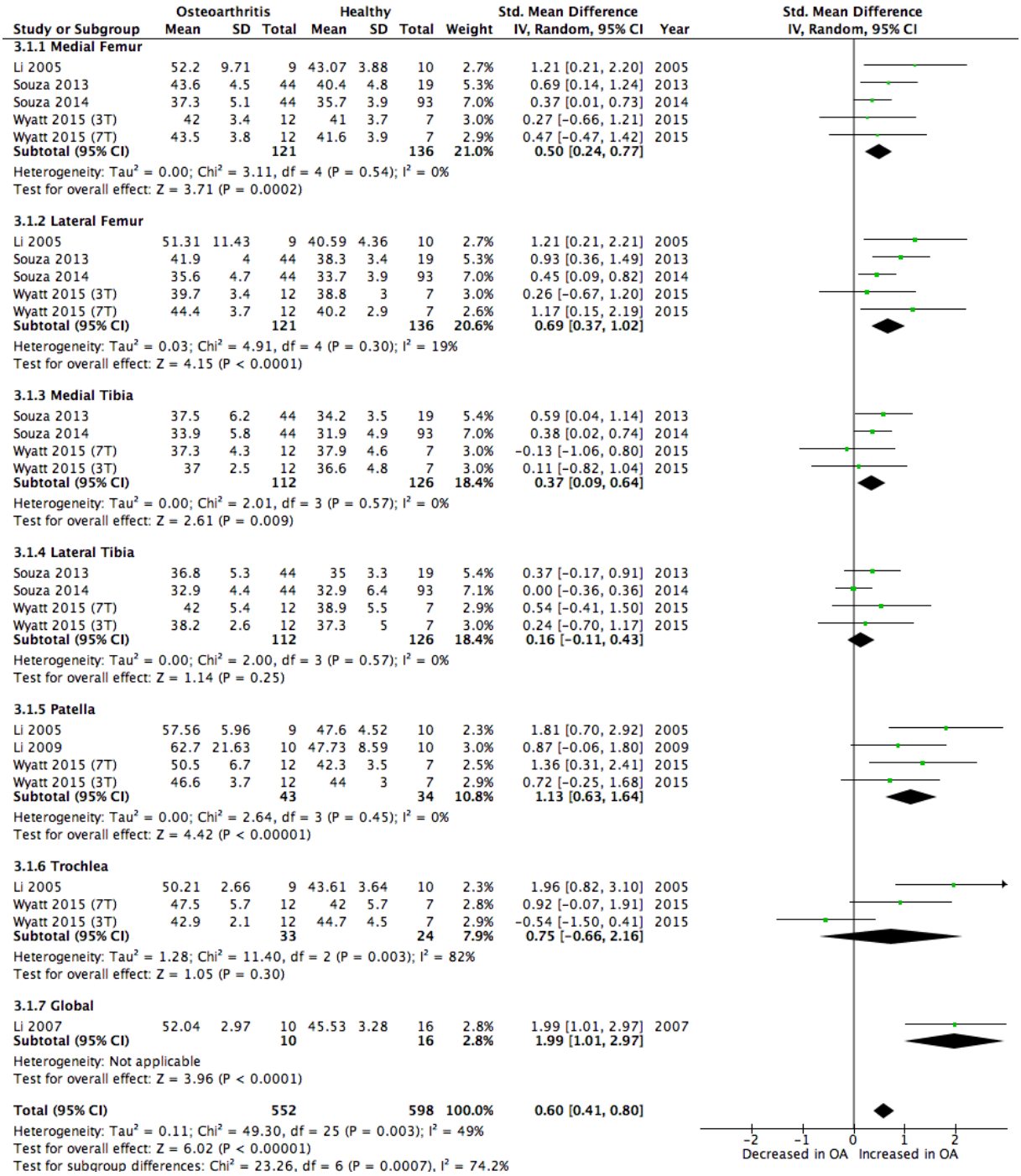
T2: OA (NOS) vs controls



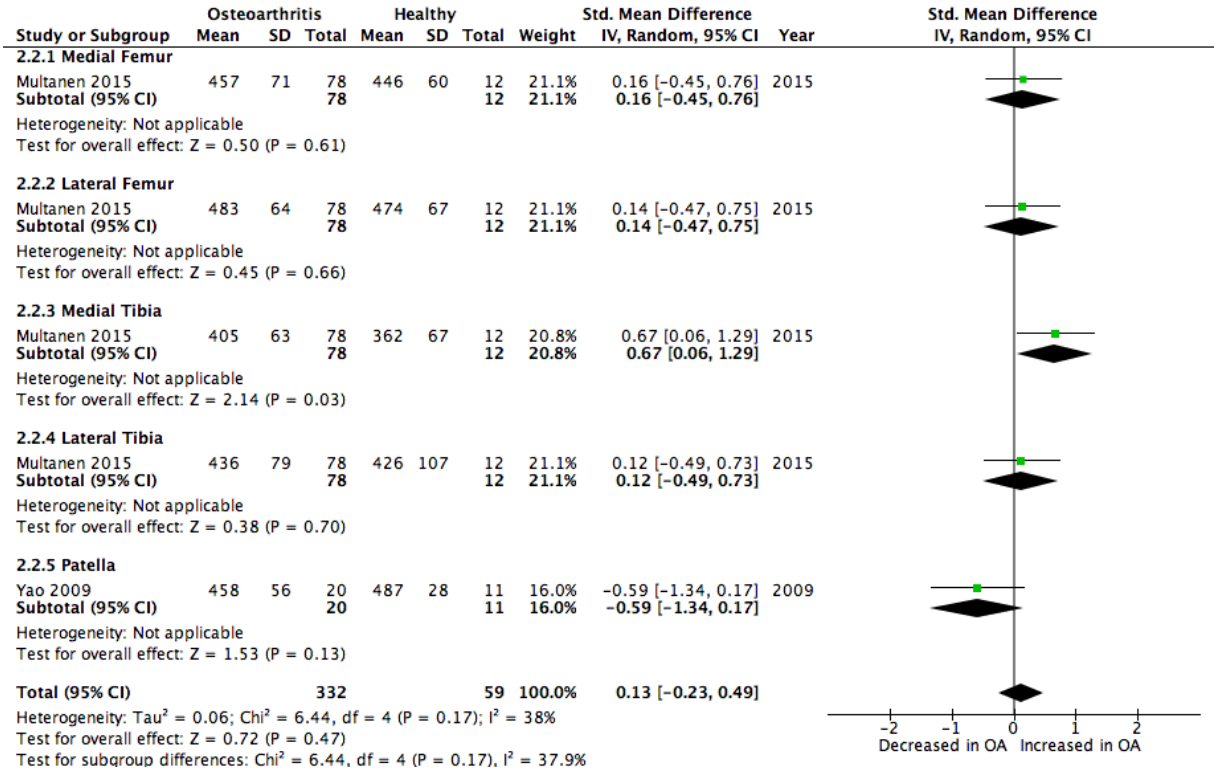
T1rho: mild OA vs controls



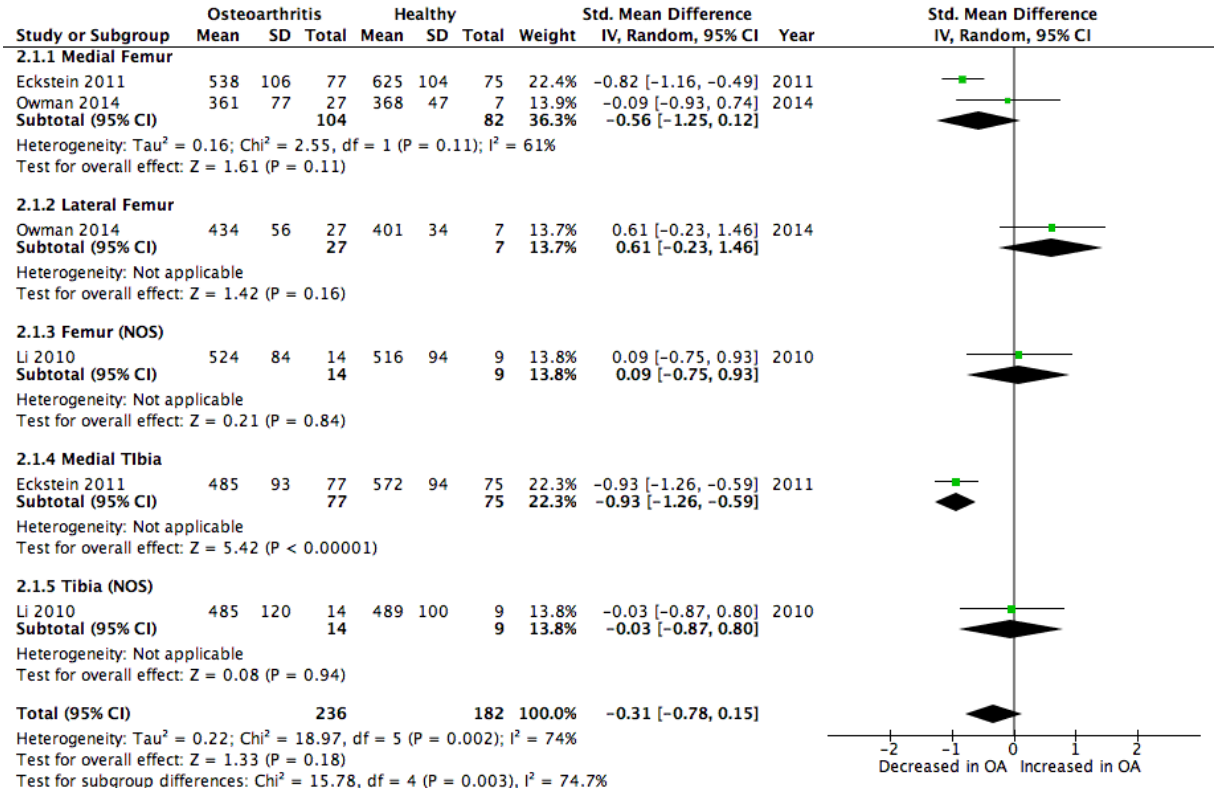
T1rho: OA (NOS) vs controls



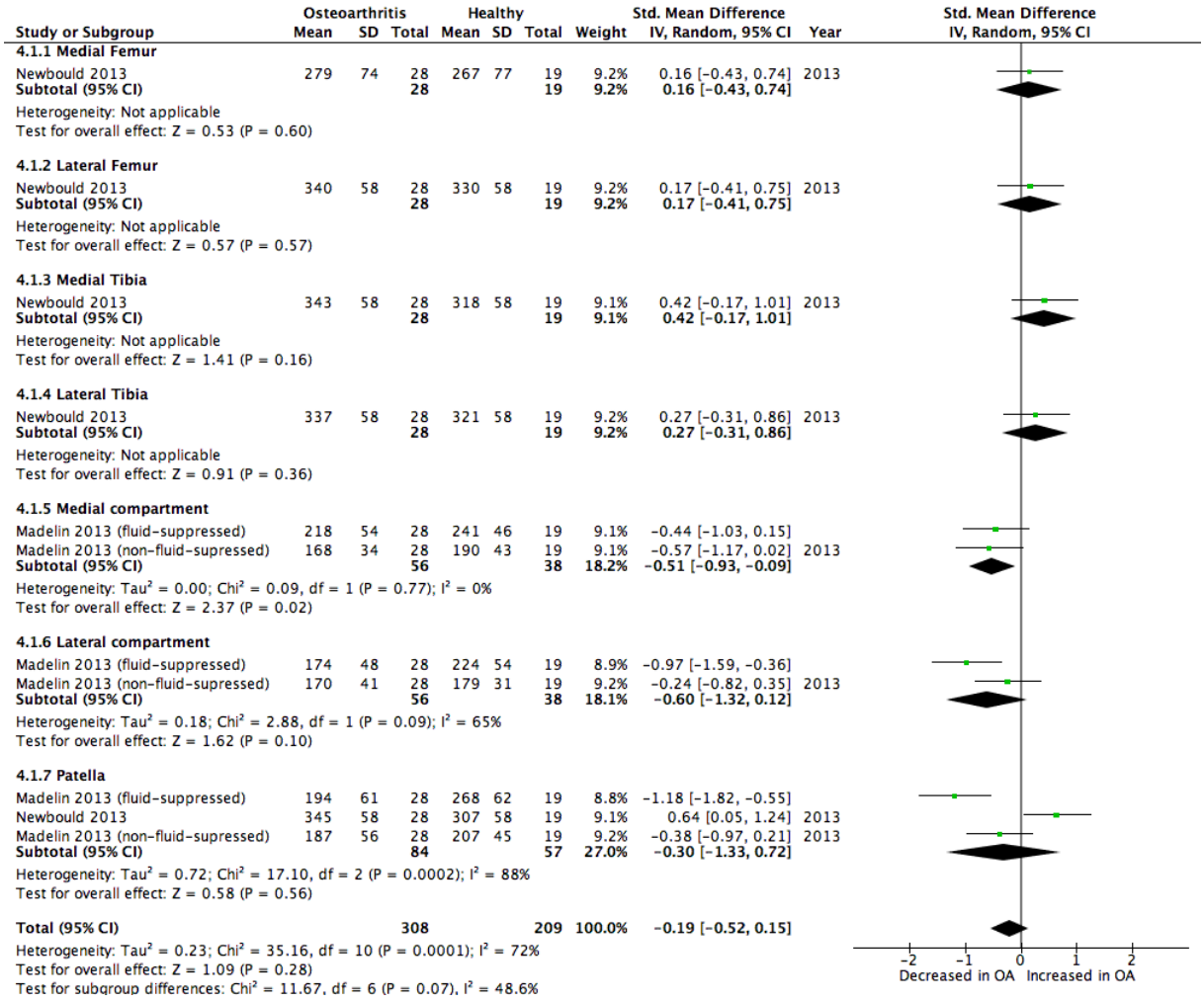
dGEMRIC: mild OA vs controls



dGEMRIC: OA (NOS) vs controls



Sodium: OA (NOS) vs controls



CHAPTER 3:

3D CARTILAGE SURFACE MAPPING – VALIDATION IN CADAVERIC KNEES

Here I describe the development of a novel semi-automatic surface-based method for analysing articular cartilage at the knee using clinical MR imaging data termed 3D Cartilage Surface Mapping (3D-CaSM). This aims to solve some of the problems associated with traditional cartilage analysis alluded to in the previous chapter. The method described allows the visualisation of the spatial distribution of cartilage thickness and compositional parameters, and the performance of spatially corresponded surface-based comparisons between individuals and within the same individual over time. I present an outline of the steps involved in the analysis pipeline and the results of a validation study using cadaveric knees.

3.1 INTRODUCTION

Improved methods of assessing articular cartilage using magnetic resonance (MR) imaging are desirable to enhance understanding of osteoarthritis (OA) onset and progression and to assess response to putative disease modifying treatments (DMOATs) targeting cartilage.

Traditional analysis of cartilage using MR data involves manual segmentation of articular cartilage surfaces. While this is accurate and reliable, it suffers from two main drawbacks. First, manual segmentation is time consuming and represents a considerable resource burden for researchers. Second, measurements are often averaged over relatively large regions of interest (ROIs). While these measurements may be attractively simple, such reductive approaches limit responsiveness and mask important focal changes(188).

Several techniques which better respect the spatial distribution of cartilage measurements have been described. However, these either rely on combining measurements made on individual 2D images which do not reflect the true 3D geometry of the cartilage surfaces, or only permit measurement of morphology and not composition(146,187,189,190).

These drawbacks limit the utility of cartilage imaging biomarkers in clinical trials, particularly in early phase clinical trials where sample sizes are small and follow-up periods are short. However, with the shift towards biomarker rich experimental

medicine study designs, there is potential for imaging to make great impact in these trials and contribute to go/no-go decision making(76).

Improved responsiveness may be realised by using better analysis methods which both reflect the true 3D spatial distribution of changes in cartilage over time and can analyse multiparametric data (i.e. both morphological and compositional). Here I propose a method which fits these criteria, termed 3D cartilage surface mapping (3D-CaSM), a modification of a previously described method termed cortical bone mapping (CBM).

CBM has been validated and used extensively for the measurement of cortical bone thickness, predominantly in the setting of osteoporosis (191,192). This method can overcome inherent inaccuracies in measurement of thin plate-like structures at clinical imaging resolutions related to slice thickness and the imaging system's point spread function via model-based deconvolution of the imaging data(193). It allows 3D measurement and visualisation of data, can handle multiparametric input and has demonstrated the ability to show targeted treatment-related effects(194–196). Visually and statistically powerful surface-based comparisons can be made between groups and within individuals over time(192). CBM-type analysis can potentially be applied to any thin plate-like structure as illustrated by its use to measure 3D joint space on CT and preliminary work demonstrating its ability to measure cartilage thickness on MR imaging(197,198).

However, the technique is yet to be validated for measuring individual cartilage surfaces at the knee. Here I aimed to validate thickness measurements obtained from cadaveric knees imaged at clinically feasible spatial resolution and acquisition times with MR imaging against very accurate high resolution peripheral quantitative CT (HRpQCT) thickness measurements obtained at a much higher spatial resolution than would be possible in vivo.

The validation of 3D-CaSM thickness values is critical for the entire 3D-CaSM pipeline, as the inner and outer cartilage surfaces generated by the thickness measurement process are crucial for downstream cartilage compositional analysis.

3.2 METHODS

3.2.1 Study subjects

Five embalmed human cadaveric knees (2 female, aged 81 – 89 years old) were obtained from the University of Cambridge Human Anatomy Centre (HAC) for the validation work. All cadavers had given ante-mortem consent for post-mortem use of specimens for education and research, and the study was approved by the University of Cambridge Human Biology Research Ethics Committee. Specimens consisted of intact articulated knee joints from distal femoral diaphysis to proximal tibial diaphysis and included all periarticular soft tissues.

3.2.2 MRI acquisition

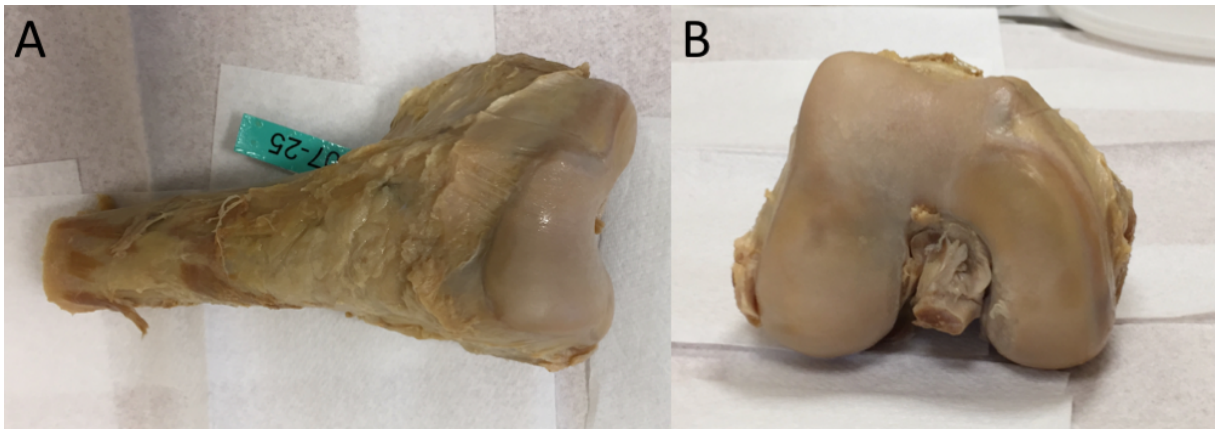
Intact cadaveric knees underwent MR imaging on a clinical 3 Tesla system (MR 750; GE Healthcare, Waukesha, WI) using an 8-channel knee coil (Invivo, Gainesville, FL). A 3-dimensional fat suppressed spoiled gradient echo (3D SPGR) sequence was performed, aiming to optimise contrast resolution between articular cartilage and adjacent bone and synovial fluid and maximise spatial resolution within a clinically feasible acquisition time. Imaging parameters were as follows: field of view, 150 x 150 x 140 mm^3; matrix size 512 x 380 (interpolated to 512 x 512) with in-plane spatial resolution 0.3 x 0.3 mm; slice thickness 1 mm; flip angle 25°; repetition time 26 ms; echo time in-phase; number of averages 0.5; acquisition time approx. 7 minutes.

3.2.3 Cadaveric dissection

Following MR imaging examination, cadaveric knees were dissected (by JM) in the HAC to allow disarticulation of the knees into the individual bones (femur, tibia and patella). The periarticular soft tissues, ligaments, capsule and menisci were removed to leave the bone with the cartilage surface exposed (Figure 3.1). To prevent desiccation, all specimens remained stored in embalming fluid when not in use.

Figure 3.1

Dissected cadaveric femur (A – frontal view, B – inferior view) showing exposure of the articular cartilage surface



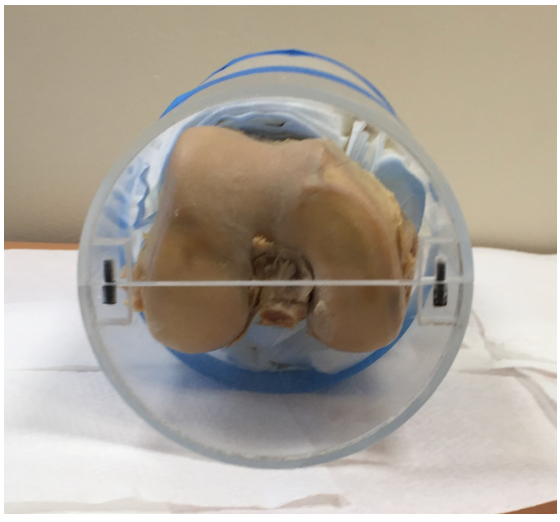
3.2.4 HRpQCT

Disarticulated specimens were mounted in an acrylic holder (Figure 3.2) and secured using padding material, ensuring that an air-cartilage interface was maintained and that the articular cartilage was not in contact with the holder at any point.

The disarticulated specimens were then examined with HRpQCT (XtremeCT; Scanco Medical, Brüttisellen, Switzerland) with an isotropic voxel size of 0.082 mm, peak voltage of 59.4 kV, tube current of 0.18 mAs.

Figure 3.2

Cadaveric femur loaded into acrylic holder for HRpQCT

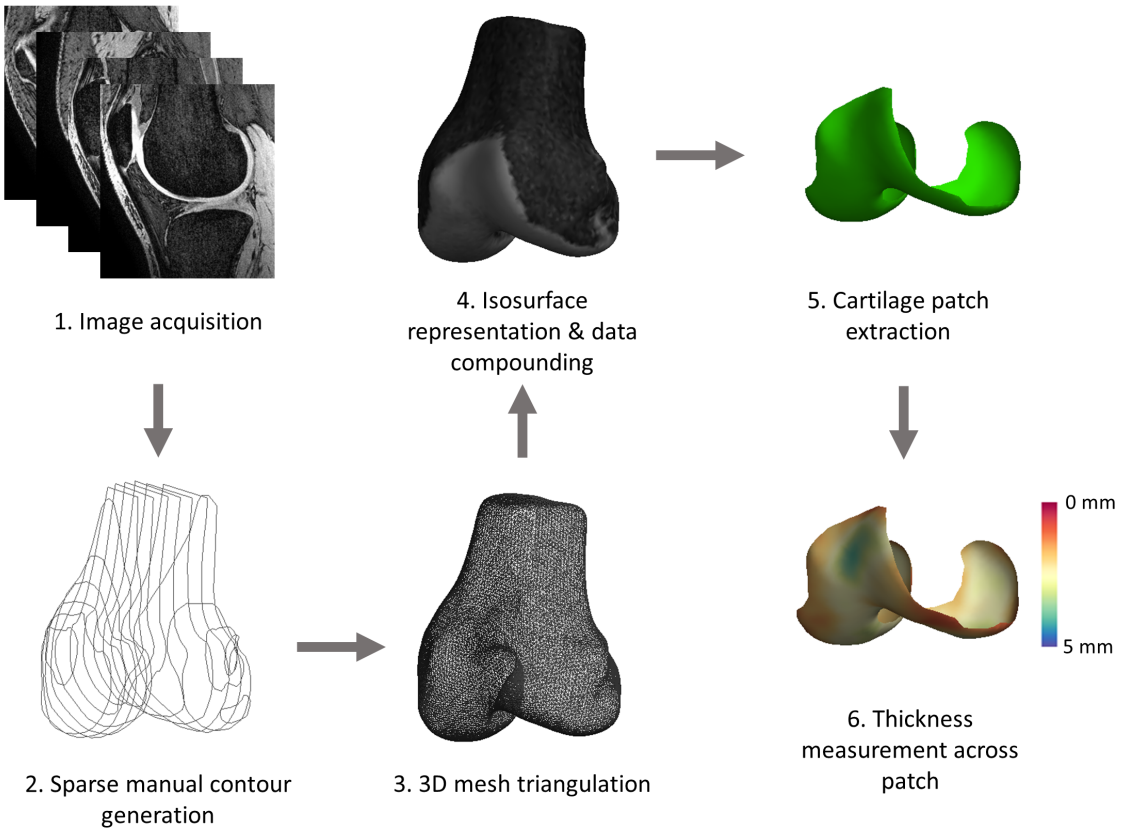


3.2.5 Image analysis

3D SPGR MR images were imported in Digital Imaging and Communications in Medicine (DICOM) format into Stradwin version 5.4a (University of Cambridge Department of Engineering, Cambridge, UK) for 3D-CaSM. The initial 3D-CaSM analysis process is summarised in Figure 3.3.

Figure 3.3

Outline of initial steps in 3D-CaSM pipeline, from image acquisition to thickness measurement. Step 6 results in the generation of accurate inner and outer cartilage surfaces between which the thickness measurements have been made. Femur used for demonstration purposes – 3D-CaSM is also performed for tibial cartilage surfaces.



First, the whole femur and tibia (including the cartilage surfaces) were manually segmented by a single observer (JM). This was performed on every 5th sagittal image (i.e. at 5 mm intervals) as the next step accurately creates the intermediate data. This step takes approximately 15 minutes per knee. Second, 3D meshes and corresponding isosurfaces were triangulated using the regularised marching tetrahedra method(199). Volume-preserving surface smoothing allows creation of an accurate surface even from sparse, approximate manual contours. This surface acts as a guide for further measurements but does not have to conform to the cartilage border exactly. Third,

cartilage 'patches' for femur, medial tibia (MT) and lateral tibia (LT) were segmented manually directly onto the 3D femoral and tibial surfaces. This is aided by the display of adjacent signal intensity values onto the 3D surface, a process known as compounding. Finally, automated cartilage thickness mapping was performed across each patch at each vertex on the surface (femoral patch - approximately 4000 vertices, medial and lateral tibial patches – approximately 1000 vertices) using a model-based deconvolution of the MR imaging data sampled along a line orthogonal to the cartilage surface (i.e. surface normal). As well as thickness measurements, this procedure also generates accurate inner and outer surfaces. This process can be prone to error in areas of low contrast resolution for the outer cartilage border (e.g. where it contacts an adjacent cartilage surface), therefore the outer surface is constrained to lie within 3 pixels (~1 mm) of the approximate manually created surface. Whole bone surfaces, inner and outer cartilage surfaces and thickness data were generated for the femur and tibia of each cadaveric knee.

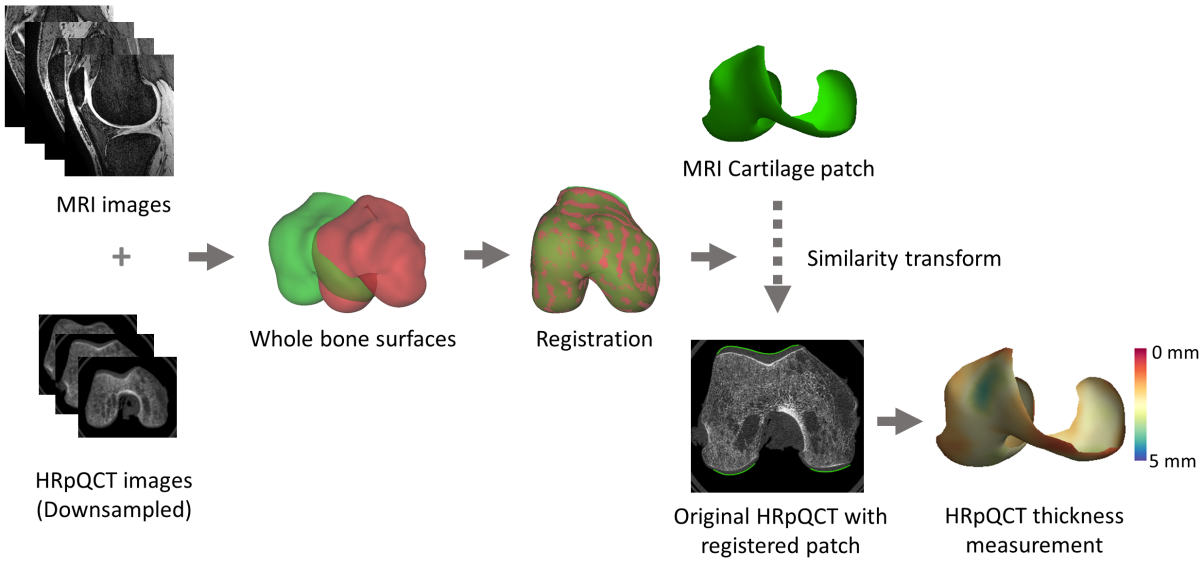
Full manual segmentation of the cartilage was also performed by the same observer. Manual thickness values were obtained by taking the distance between the intersections of the outer and inner surface on the manual segmentation with a line normal to each vertex on the corresponding cartilage patch generated by 3D-CaSM, ensuring vertexwise correspondence of 3D-CaSM and manual thickness values.

HRpQCT images were downsampled to 1 mm isotropic resolution and smoothed with a 3D Gaussian filter to facilitate initial analysis. Surfaces were created for the whole femur and tibia from downsampled HRpQCT data. The MR surfaces were then registered to the HRpQCT surface for the corresponding whole bone with dedicated surface registration software (wxRegSurf version 18; University of Cambridge Department of Engineering, Cambridge, UK), using a similarity transform. The transformation matrix for registering the MR to the HRpQCT whole bone surface was then applied to the corresponding cartilage surface extracted from the MR data. The original (i.e. neither downsampled nor blurred) HRpQCT data and transformed cartilage surfaces were then loaded into Stradwin. The transformed surface was used as a guide for cartilage thickness measurement of the HRpQCT data as described above, meaning that the MR–HRpQCT registration did not have to be exact. Due to contrast differences between MR and HRpQCT and the fact that pixel (Hounsfield

unit) values on HRpQCT are (unlike MR) not arbitrary, automated thickness measurements on the HRpQCT data were performed using a threshold-based full-width half-maximum (FWHM) technique (Figure 3.4).

Figure 3.4

Outline of HRpQCT thickness measurement process. This results in a set of thickness measurements at each HRpQCT cartilage surface vertex which correspond to the measurement at the same vertex on the MR surface. Femur used for demonstration purposes, same process performed for tibial data.



3.2.6 Statistical analysis

The thickness value at each vertex on each cadaveric MR surface was compared to the thickness value at the identical location on the corresponding HRpQCT surface for both 3D-CaSM and manual segmentation. Correlation between the two measurements was assessed visually using scatter plots. Bland-Altman analysis was performed to calculate the mean bias and 95% limits of agreement between MR and HRpQCT thickness measurements. The root-mean-square-average-error (RMSE) was also calculated to facilitate comparison with the work of others. These analyses were performed separately for each cartilage surface and also with data combined from all surfaces. By performing appropriate spatial normalisation (surface-to-surface similarity and thin-plate spline registration performed in wxRegSurf), the spatial distribution of MR-HRpQCT error values across all cadaveric subjects could be displayed on single

representative surfaces. Statistical analyses were performed using RStudio version 1.0.143.

3.3 RESULTS

3.3.1 MR imaging

All five cadaveric knees were successfully imaged with MR. The measurement technique in Stradwin works well and is simple to perform, taking approximately 20 minutes per knee to analyse all cartilage surfaces. Automated cartilage edge detection agrees subjectively with visual assessment of cartilage boundaries.

3.3.2 HRpQCT validation

Data from 4 femoral and 4 tibial specimens were analysed (one cadaveric knee was excluded due to presence of end-stage OA with very little residual cartilage). This gave a total of 17,335 surface vertices for comparison between the two methods.

Table 3.1

Results of Bland-Altman analyses for comparison of MR and HRpQCT thickness values. **Abbreviations:** RMSE – root mean square error, MT – medial tibia, LT – lateral tibia.

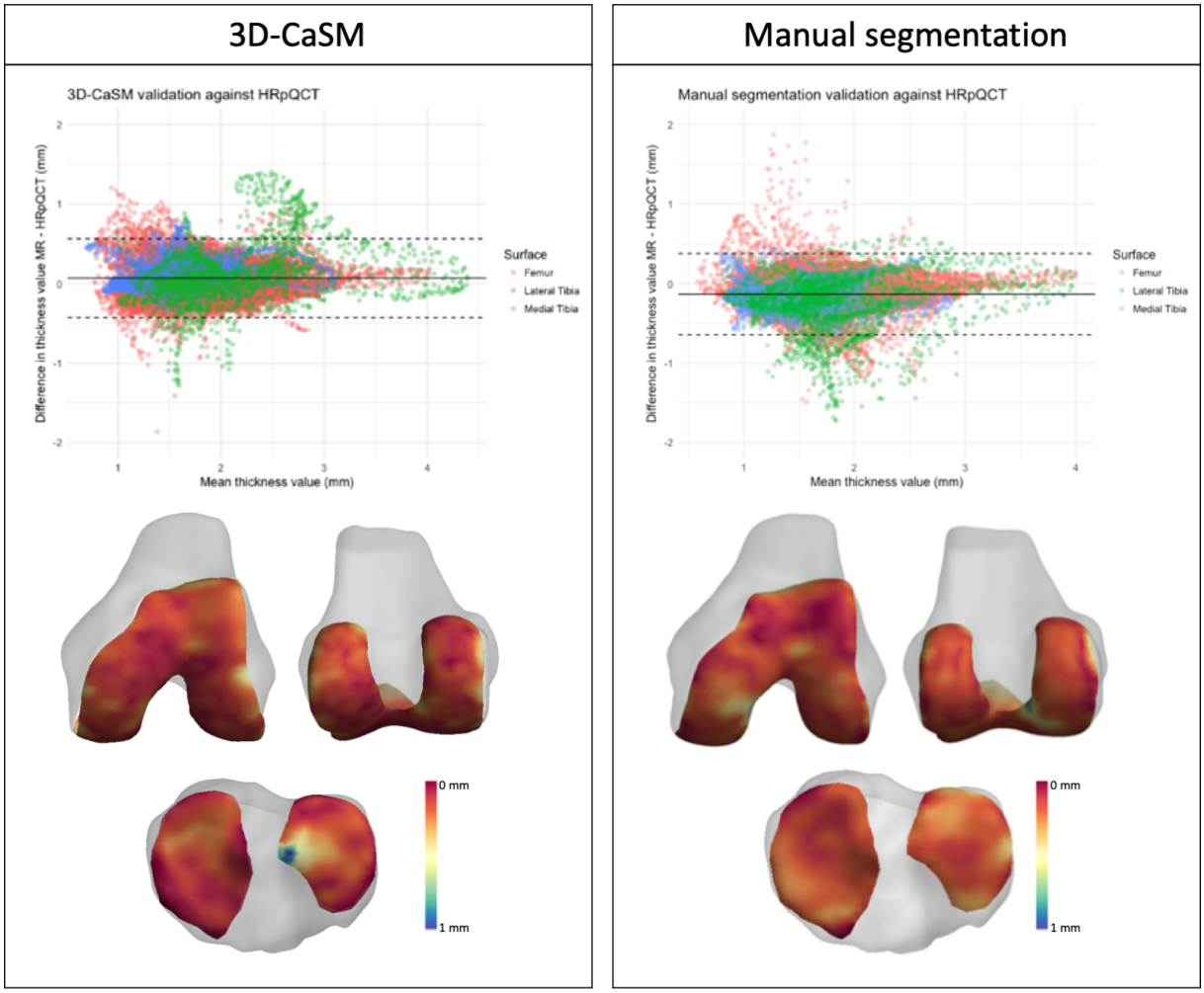
Surface	Method	Mean bias MR – HRpQCT (mm)	95% limits of agreement (mm)	RMSE (mm)
Femur	3D-CaSM	0.05	-0.40, 0.50	0.23
	Manual	-0.1	-0.58, 0.38	0.26
MT	3D-CaSM	0.15	-0.21,0.52	0.24
	Manual	-0.17	-0.53, 0.19	0.25
LT	3D-CaSM	0.11	-0.64, 0.85	0.40
	Manual	-0.25	-0.95,0.45	0.44
All (combined)	3D-CaSM	0.06	-0.43,0.56	0.26
	Manual	-0.13	-0.64,0.38	0.29

Cadaveric validation data are provided in Table 3.1. 3D-CaSM MR thickness measurements demonstrated mean bias (95% limits of agreement) of +0.05 mm (-0.40 to 0.50 mm), +0.15 mm (-0.21 to 0.52 mm) and +0.11 mm (-0.64 to 0.85 mm) for femur, MT and LT respectively when compared to the reference HRpQCT data. The spatial distribution of MR errors was fairly uniform across the cartilage surfaces with

the exception of the extreme medial portion of the LT (Figure 3.5). Mean bias and 95% limits of agreement between manual MR thickness measurements and HRpQCT were -0.1 mm (-0.58 to 0.38 mm), -0.17 mm (-0.53 to 0.19 mm) and -0.25 mm (-0.95 to 0.45 mm) for femur, MT and LT respectively (Figure 3.5).

Figure 3.5

Comparison of agreement between 3D-CaSM and HRpQCT and manual segmentation and HRpQCT. Upper panels - Bland-Altman plots. Solid lines represent mean bias, dotted lines represent 95% limits of agreement. Lower panels - average error for all cadaver knees displayed on canonical surfaces.



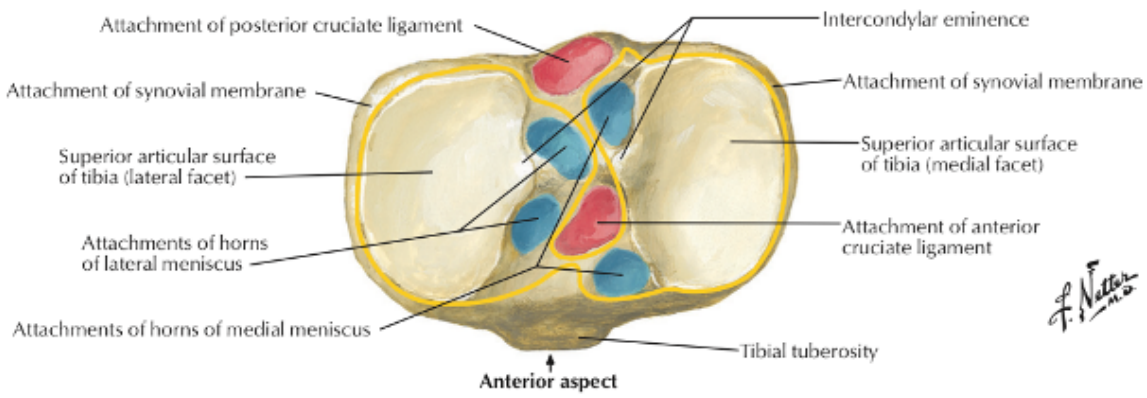
3.4 DISCUSSION

This study provides evidence for the construct validity of 3D-CaSM. Bias and precision of 3D-CaSM are comparable to that of manual segmentation which is often regarded as the 'gold standard', but with the added benefits of significant reduction in time taken for analysis and the possibility of performing surface-based analyses which may allow us to learn more from the data than is obtained from averaging measurements over large ROIs.

For all surfaces analysed, mean bias of 3D-CaSM was ≤ 0.15 mm with 95% limits of agreement less than the largest edge of a voxel in the images analysed (1 mm), i.e. sub-voxel accuracy. The only region with errors that could be considered as unacceptable was the extreme medial part of the lateral tibia where the cartilage surface slopes caudally to meet the lateral intercondylar eminence, close to the lateral meniscal roots (Figure 3.6). During dissection, it was difficult to remove the deepest parts of the meniscal roots without risking damage to the adjacent cartilage surface (which would have invalidated the experiment), therefore a relatively conservative dissection of these regions was performed. It is therefore hypothesised that the poor performance of 3D-CaSM vs HRpQCT in this region may be due to residual (undissected) ligamentous/fibrous tissue confounding measurement.

Figure 3.6

Schematic illustration of the tibial plateau (superior view) demonstrating major ligamentous and meniscal attachment sites. Reproduced from (200).



The results demonstrate a tendency for 3D-CaSM to over-estimate cartilage thickness by a small amount and for manual segmentation to under-estimate thickness by a small amount when compared to the reference method. This is probably because

experienced manual segmenters will tend to be cautious when defining the outer cartilage edge to avoid partial volume effects, whereas 3D-CaSM relies on a signal model meaning that voxels adjacent to the outer cartilage edge with similar signal intensity will tend to be included, resulting in a small over-estimation. This phenomenon has been observed previously with other automated segmentation methods(201,202).

In this study I have used HRpQCT as a reference method in preference to the more commonly used expert manual segmentation in the belief that manual segmentation, even if performed by an expert, is prone to observer error and thus is less than ideal as a reference method. The result of a small but systematic under-estimation of cartilage thickness using expert manual segmentation supports this approach and should be considered by others when designing similar validation experiments.

Obtaining comparable agreement statistics for previously published methods is difficult due to the variety of statistical methods used for assessing agreement and the different validation methods used. However, bias and precision are broadly similar to previously described methods. Methods with narrower limits of agreement with a manual segmentation reference method have been described, for example a popular active appearance model-based method, but these limits of agreement have been derived from measures averaged across large ROIs rather than from individual measurement points (203). In addition, it should be borne in mind that 3D-CaSM allows analysis of not only cartilage thickness but also (given the appropriate input) cartilage composition, something not achieved by existing methods.

The clinical importance of 3D-CaSM is illustrated by the achievements of CBM in the field of osteoporosis, where the method has provided new insights into the efficacy of osteoporosis therapies, interaction between therapies and risk factors for disease progression and fracture. While the development of 3D-CaSM is at an earlier stage, it offers the possibility of similar achievements in OA. 3D-CaSM also opens up the possibility of novel surface-based analysis of cartilage data. In larger datasets, this should be performed using statistical parametric mapping (SPM), a well-established method in the neuroimaging field which combines linear mixed-effect models with Gaussian random field theory to correct for multiple comparisons. In smaller datasets, such as in experimental medicine studies which are the focus of this thesis, SPM may

be less suitable due to the small sample size. In this setting, alternative descriptive analytical approaches, such as applying a threshold considered meaningful to changes over time, may be suitable. The opportunity that 3D-CaSM offers to detect and localise even very focal changes in cartilage morphology or composition may allow early assessment of treatment efficacy not accessible to alternative approaches.

There are several limitations of the current work. First, this validation study featured only four cadaveric knees. The number of knees available for this study was limited due to the fact that dissection was required, limiting the utility of residual cadaveric material for future educational use. However, these four knees generated over 17,000 data points for validation across a clinically meaningful range of cartilage thickness values and would be considered ample by conventional sample size calculation. Second, one cadaveric knee was excluded due to the presence of end-stage OA with extensive full thickness cartilage loss. While the utility of this cadaveric knee for validation would have been limited, it should be noted that 3D-CaSM may not be valid for measurement in end-stage disease. Individuals with such severe OA are highly unlikely to be selected for trials of potential DMOATs therefore the lack of utility of 3D-CaSM in this setting is not of undue concern. Finally, using HRpQCT as a reference method required dissection of the knees with removal of any tissues overlying the cartilage which could confound measurement yet avoidance of damage or removal of any of the cartilage itself. As discussed above, this was challenging at entheseal sites close to the cartilage surface such as the meniscal horns, cruciate ligaments, gastrocnemius tendons meaning that there is some risk that this may have confounded reference measurement. However, the vast majority of measurement points used for validation were not affected by this problem. Moreover, there are similar problems at the edges of the cartilage surfaces with manual segmentation which has not precluded its use as a reference method in previous validation studies.

In conclusion, this study provides evidence of the construct validity of 3D-CaSM. The validation of thickness measurement by 3D-CaSM is important for the downstream sampling of compositional data which relies on the accuracy of the inner and outer cartilage surfaces generated by the thickness measurement process. The initial clinical application of this work is described in chapter 8.

CHAPTER 4:

ASSOCIATION OF SUBCHONDRAL BONE TEXTURE ON MAGNETIC RESONANCE IMAGING WITH RADIOGRAPHIC KNEE OSTEOARTHRITIS PROGRESSION – DATA FROM THE OSTEOARTHRITIS INITIATIVE BONE ANCILLARY STUDY

MacKay, J. W. et al. Association of subchondral bone texture on magnetic resonance imaging with radiographic knee osteoarthritis progression: data from the Osteoarthritis Initiative Bone Ancillary Study. Eur Radiol 28, 4687–4695 (2018).

Subchondral bone is increasingly thought to be at least as important as articular cartilage in OA pathogenesis. However, the development of quantitative imaging biomarkers of subchondral bone on imaging has lagged behind the development of those for cartilage. In this chapter, I use the technique of texture analysis for subchondral bone assessment. The promise of texture analysis in this setting is in that it will detect alterations in trabecular architecture occurring in OA which manifest as changes in the heterogeneity and spatial organisation (i.e. 'texture') of the subchondral bone on an image.

This chapter involves analysis of 122 participants in the Osteoarthritis Initiative, a prospective multicentre observational study. I assess the association of subchondral bone texture on MR imaging with the radiographic worsening of OA over a 3-year period, as well as the ability of subchondral bone texture to predict this worsening. My findings show that subchondral bone texture is associated with OA progression, and that analysis of subchondral bone texture can discriminate between individuals with OA who progress vs those who do not.

4.1 INTRODUCTION

It is increasingly recognised that subchondral bone plays a critical role in osteoarthritis (OA) onset and progression. Subchondral bone is a dynamic tissue, absorbing the majority of forces transmitted through the joint and capable of remodelling in response to stress(204). Therefore, there has been increasing interest in subchondral bone as a target for potential disease modifying OA treatments (DMOATs)(205).

Sensitive markers of subchondral bone alterations that occur in OA are required for such treatments to be evaluated. Several imaging biomarkers of subchondral bone have been described using plain radiographs, dual x-ray absorptiometry (DXA), computed tomography (CT) and magnetic resonance (MR) imaging, including direct estimation of trabecular microarchitecture and fractal signature analysis (FSA) (206,207). While several established biomarkers have shown cross-sectional associations with OA severity, there remains room for improvement with regard to the ability to predict OA progression(40,208–210).

MR texture analysis (MR TA) has recently been described as a method of quantifying subchondral bone changes in OA which may offer superiority over existing biomarkers(211). MR TA involves the calculation of several statistical descriptors of image texture, aiming to characterise the heterogeneity and spatial organisation of the subchondral bone. The technique has been shown to be reproducible and able to distinguish subjects at different stages of OA from healthy subjects(211). Moreover, texture features have shown an improved ability to discriminate between knees with and without OA when compared to microarchitectural analysis and are significantly associated with histomorphometry(212,213).

However, MR TA has previously only been used to compare subjects with OA and healthy controls. To be useful as a prognostic or treatment evaluation imaging biomarker, it should also be able to identify which individuals with OA are likely to progress (predictive validity) and demonstrate sensitivity to change for OA progression (concurrent validity).

Therefore, the purpose of this study was to evaluate whether initial (i.e. OAI visit at 30 or 36-months) or 12-18 month change (i.e. between OAI visit at 30/36-months and 48-months) in MR subchondral bone texture were predictive of radiographic OA progression over 36-months (i.e. between OAI visit at 36-months and 72-months).

4.2 METHODS

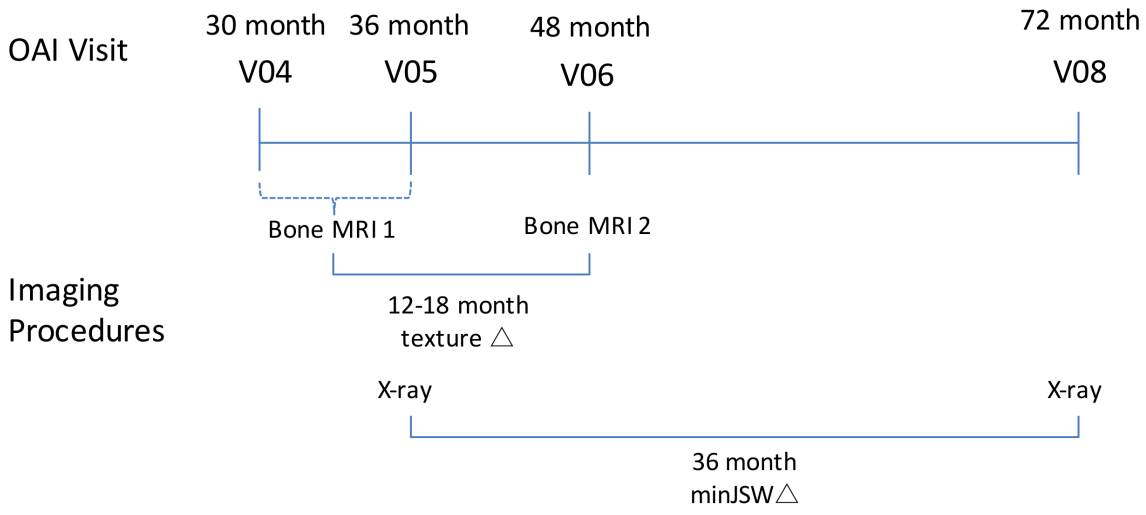
The Osteoarthritis Initiative (OAI) has been approved by the institutional review boards for the University of California, San Francisco and the four OAI clinical centres (University of Pittsburgh, Ohio State University, University of Maryland, Baltimore, and Memorial Hospital of Rhode Island). All participants have given informed consent to participate in the study. This was a retrospective nested case-control study. The OAI datasets are freely available for download at https://oai.epi-ucsf.org.

4.2.1 Participants

Included participants were participants in the OAI bone ancillary study (BAS). The BAS featured 629 participants who underwent MR examination optimised for assessment of subchondral bone in addition to the standard OAI MR sequences. All BAS participants were members of the progression subcohort, participants with both frequent knee symptoms and radiographic OA in at least one knee at OAI inception. Initial trabecular bone MRs were performed at either the 30 or 36-month OAI visit, with the majority of participants undergoing a repeat MR at the 48-month OAI visit ("12-18 month follow-up"). Study design is summarised in Figure 4.1.

Figure 4.1

Study timeline for image assessments as part of the Osteoarthritis Initiative Bone Ancillary Study.



Study cases ("progressors") were defined as individuals with radiographic progression over a 36-month period between the 36-month OAI visit and the 72-month OAI visit, according to the definition of the Foundation for the National Institutes of Health (FNIH) OA biomarkers consortium: a decrease in minimum medial tibiofemoral joint space width (minJSW) of ≥ 0.7 mm, which has a less than 10% chance of being due to measurement error(214). The details of the radiographic acquisition and assessments have been discussed in detail previously. Briefly, knee radiographs were performed using a non-fluoroscopic fixed flexion technique(215). Assessments included central readings for Kellgren-Lawrence (KL) grading, Osteoarthritis Research Society International (OARSI) grading of joint space narrowing (JSN), measurement of femorotibial alignment and automated measurement of joint space width(216–218).

Progressors were matched to participants who did not have radiographic progression between the 36-month OAI visit and 72-month OAI visit (controls) for age, sex, body mass index (BMI) and initial medial minJSW in a 1:1 ratio, using an optimal nearest-neighbour propensity score algorithm. Although the OAI featured a healthy reference subcohort, none of this group underwent dedicated trabecular bone MR imaging and therefore were not available to use as a control group in this study.

Individuals who did not have measurements of minJSW available at the 36-month or 72-month OAI visits were excluded, as were individuals with KL grade 4 knees at the 36-month visit (due to ceiling effects on minJSW progression in this group). As this study focused on the medial tibiofemoral compartment, any individuals who had lateral compartment predominant disease at the 36-month OAI visit were also excluded, as defined by greater OARSI JSN grade in the lateral than medial compartment.

4.2.2 MR Acquisition

Study participants were evaluated with a coronal-oblique 3-dimensional fast imaging with steady state precession (FISP) MR sequence (field of view 12 x 12 cm, matrix 512 x 512 (interpolated to 1024 x 1024), slice thickness 1 mm, repetition time 20 ms, echo time 4.92 ms, flip angle 50°, number of signal averages 1, acquisition time 10.5 minutes) optimised for visualisation of subchondral trabecular bone(219). This was

performed on one of four identical Siemens Trio 3T MR platforms used for the OAI using a quadrature transmit-receive knee coil (USA instruments).

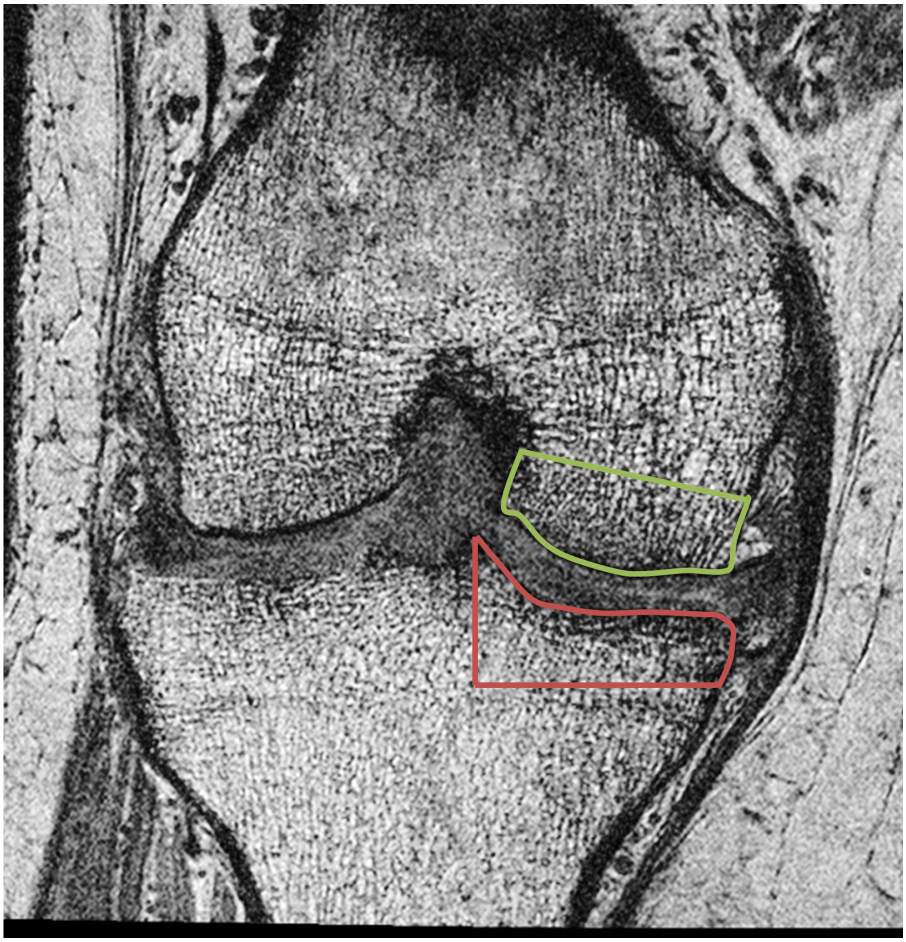
4.2.3 MR Analysis

The five most central coronal-oblique images through the central medial tibiofemoral joint were identified with reference to axial and sagittal reformats and used for subsequent analysis.

The MR images were imported into a dedicated texture analysis program (MazDA v3.3, freely available at http://www.eletel.p.lodz.pl/programy/mazda/)(220). Regions of interest (ROIs) were created manually in the medial tibial and medial femoral subchondral bone on each coronal image with blinding to case or control status of participants. ROIs were defined superiorly and inferiorly by the bone-cartilage interface, medially and laterally by the margins of the tibial plateau and femoral condyle and extended for a depth of approximately 1 cm into the subchondral bone, ensuring that no part of the physeal plate was included. Illustrative ROI examples are provided in Figure 4.2.

Figure 4.2

Example coronal-oblique 3D FISP MR image demonstrating ROI placement in the medial tibial (red) and medial femoral (green) subchondral bone.



Nineteen texture features (listed in Appendix 4.1) were calculated for each ROI aiming to quantify the heterogeneity and spatial organisation of the subchondral bone, according to the method described previously(211). These texture features belonged to one of four classes: grey-level histogram, absolute gradient, run-length matrix (RLM) and grey level co-occurrence matrix (GLCM). Briefly, grey-level histogram features are simple descriptors of the distribution of grey levels (i.e. pixel intensity values) in the ROI. Gradient, RLM and GLCM features are higher order descriptors of the spatial organisation of pixels in the ROI. Image compression settings of 4 bits/pixel for calculation of gradient features, and 6 bits/pixel for calculation of GLCM and RLM parameters were used. RLM parameters were calculated four times for each pixel (in the horizontal, vertical, 45° and 135° directions) and GLCM parameters were calculated 20 times for each pixel at a variety of pixel offsets ranging from 1 to 5 pixels. The mean

value of each RLM and GLCM parameter for each pixel in all possible directions and pixel offsets was calculated for each coronal image. The values of each texture parameter on each of the five coronal images analysed were then averaged to give summary values in each participant for medial tibial and medial femoral ROIs.

4.2.4 Texture analysis reproducibility

Twenty-three participants were randomly selected using a random number generator (www.random.org) and analysed in duplicate by two independent analysts to assess reproducibility. Analysts created ROIs independently and were blinded to case or control status. The sample size was based on previous data suggesting a mean intraclass correlation coefficient (ICC) value of 0.9 across texture features(212,221). Texture features with suboptimal reproducibility metrics (ICCs of < 0.8 or root-mean-square average coefficient of variation (RMSCV) of > 10% for either ROI) were excluded from subsequent analyses.

4.2.5 Statistical analysis

Descriptive statistics for each texture feature were generated. The distribution of texture values in progressors and controls was compared visually using boxplots. Composite texture scores using linear combinations of texture features were created using least absolute shrinkage and selection operator (LASSO) penalised logistic regression. This helps to avoid problems associated with overfitting when many predictor variables are available by imposing a penalty for including additional variables in a model. The LASSO model was chosen based on minimisation of 10-fold cross-validation classification error. Because the folds for cross-validation are chosen at random, the procedure was repeated with 100 iterations. The logistic regression equation from a model iteration giving the mean classification error was used to calculate a composite texture score. This was performed separately for tibial, femoral and combined (tibial and femoral) texture datasets using initial MR texture features and 12-18 month change in MR texture features.

Odds ratios were calculated for the increase or decrease in odds of radiographic progression per one standard deviation (SD) increase in texture score. The mean classification accuracy (*c*-statistic, equivalent to area under the receiver operating curve (AUC)) was recorded.

Inter-observer reproducibility was assessed using ICC values (single measures, absolute agreement) and the RMSCV for each texture feature.

All statistical analyses were performed using RStudio version 1.0.136 for Mac, using the *MatchIt* package for matching cases to controls, and the *glmnet* package for performing LASSO regression(222,223). Statistical significance of the logistic regression analyses was assessed using the chi-squared test, with an adjusted *P* value threshold of < 0.008 to maintain an overall type 1 error rate of 0.05.

4.3 RESULTS

4.3.1 Participants

Table 4.1

Subject characteristics at initial timepoint and 12-18 month follow-up

Variable	Initial		12-18 month follow-up	
	Cases n = 61	Controls n = 61	Cases n = 53	Controls n = 52
Age, years*	64 (49 – 81)	65 (48 – 82)	65 (50 – 82)	66 (49 – 83)
Sex, no. females	25	26	21	22
BMI, kg/m^2 **	31.4 (4.7)	31.1 (4.7)	31.1 (4.5)	30.8 (4.6)
Time between baseline and follow-up MRI (12/18 months)	-	-	35/18	26/26
Initial minJSW, mm**	3.81 (1.20)	3.78 (1.19)	3.86 (1.19)	3.75 (1.20)
Kellgren Lawrence grade (0/1/2/3)	7/11/21/22	8/9/24/20	7/11/17/18	8/9/20/15
OARSI medial JSN grade (0/1/2)	17/22/22	23/18/20	15/20/18	21/16/15
OARSI lateral JSN grade (0/1/2)	60/1/0	58/3/0	52/1/0	50/2/0
Femorotibial alignment, degrees†	-6.0 (1.9)	-5.7 (2.1)	-6.0 (1.8)	-5.7(2.1)
minJSW change, mm**	-	-	-1.29 (0.63)	0 (0.44)

*mean (range)
**mean (standard deviation)
†mean (standard deviation), negative values indicate varus alignment

Of 629 participants in the BAS, 359 were eligible for this study following exclusions. 64 participants met the criterion for radiographic progression and were selected as cases, with 64 controls matched for age, sex, BMI and initial minJSW.

For initial timepoint analyses, 12 participants (3 cases) were excluded due to excessive motion artefact, defined as preventing the identification of the bone-cartilage interface (n = 5), unavailable initial MR (n = 3) or large subchondral cysts (n = 4). The matched controls for excluded cases were also removed from the analysis. For each excluded control, a new matched control was selected as the next best match for the corresponding case according to the matching algorithm.

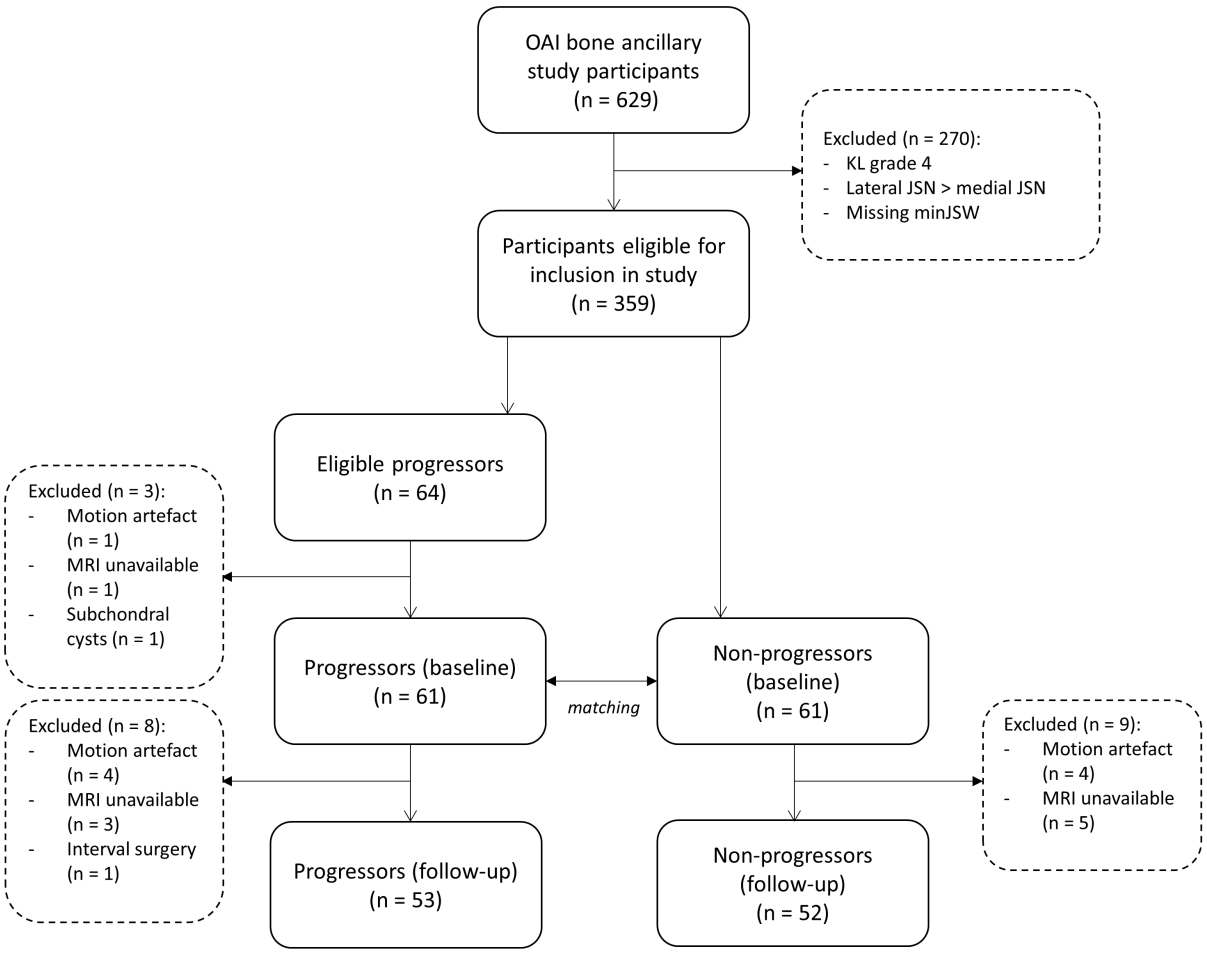
For 12-18 month follow-up analyses, a further 17 participants were excluded due to excessive motion artefact on follow-up images (n = 8), unavailable follow-up MR (n =

8), and surgical intervention involving the subchondral bone during the interval between initial and follow-up MR (n = 1).

Initial and follow-up characteristics of included cases and controls are presented in Table 4.1. A flow diagram for selection of study participants is presented in Figure 4.3.

Figure 4.3

Flow diagram for selection of study participants



4.3.2 Texture analysis reproducibility

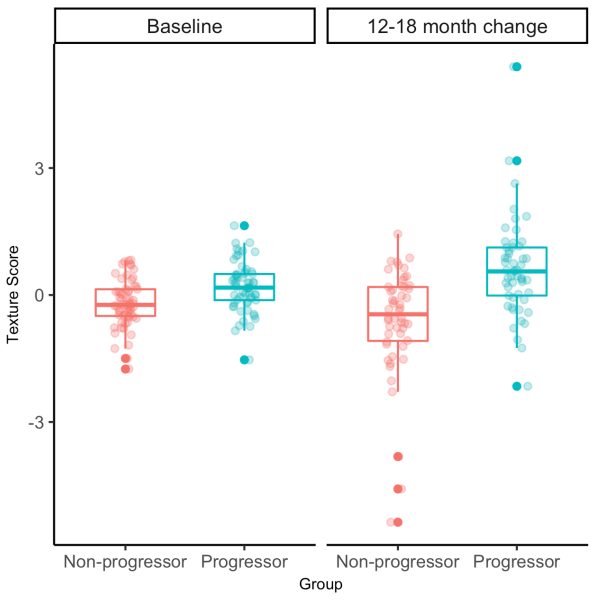
Data for inter-observer reproducibility are presented in Appendix 4.1. The majority of texture features demonstrated excellent inter-observer reproducibility. Seven texture features were excluded from subsequent analyses at this stage (per criteria in section 4.2.4), leaving a total of 12 texture features for analysis.

4.3.3 Association of subchondral bone texture and radiographic progression

Subchondral bone composite texture score was significantly associated with 36-month radiographic progression using the initial timepoint combined data (odds ratio [95% confidence interval] = 1.84 [1.25 – 2.80], $P = 0.002$) and the 12-18 month change tibial (2.31 [1.42 – 4.12], $P < 0.001$), femoral (1.80 [1.17 – 2.92], $P = 0.006$) and combined (3.76 [2.04 – 7.82], $P < 0.001$) data. Associations between subchondral bone texture score and radiographic progression using initial timepoint tibial (1.43 [0.99 – 2.09], $P = 0.06$) and femoral (1.63 [1.12 – 2.44], $P = 0.009$) data were not statistically significant. Results are summarised in Table 4.2 and Figure 4.4. Data for each individual texture feature are provided in Appendix 4.2. Example images are shown in Figure 4.5.

Figure 4.4

Boxplot demonstrating the association between subchondral bone composite texture score and radiographic progression status using baseline and 12-18 month change combined tibial and femoral data. Individual datapoints are displayed as semi-transparent dots.



4.3.4 Texture analysis classification

Combinations of both initial texture features and 12-18 month change in texture features were able to predict radiographic progression with statistical significance reached for femoral and combined initial data, and tibial, femoral and combined 12-18 month change data. The best classification accuracy was demonstrated for combined 12-18 month follow-up data with a *c* statistic of 0.68 (95% confidence interval (CI) 0.68 – 0.68, $P < 0.001$).

Classification performance is summarised in Table 4.2.

Table 4.2

Association between texture features and case vs control status, and classification performance

Region	Initial			12-18 month change		
	Odds ratio (95% CI)†	Most important features#	*c*-statistic (95% CI)	Odds ratio (95% CI)†	Most important features#	*c*-statistic (95% CI)
Tibia	1.43 (0.99, 2.09)	GrVar Variance	0.58 (0.58, 0.58)	2.31 (1.42, 4.12)***	GrMean GrVar Contrast	0.65 (0.63, 0.69)**
Femur	1.62 (1.12, 2.44)**	Mean Variance ASM	0.60 (0.59, 0.60)*	1.80 (1.17, 2.92)**	ASM Contrast Variance	0.63 (0.61, 0.65)**
Combined	1.84 (1.25, 2.80)**	F Variance F Mean T Variance	0.64 (0.64, 0.65)**	3.76 (2.04, 7.82)***	T GrVar F ASM T Entropy	0.68 (0.68, 0.68)***

* P < 0.05, ** P < 0.01, ***P < 0.001

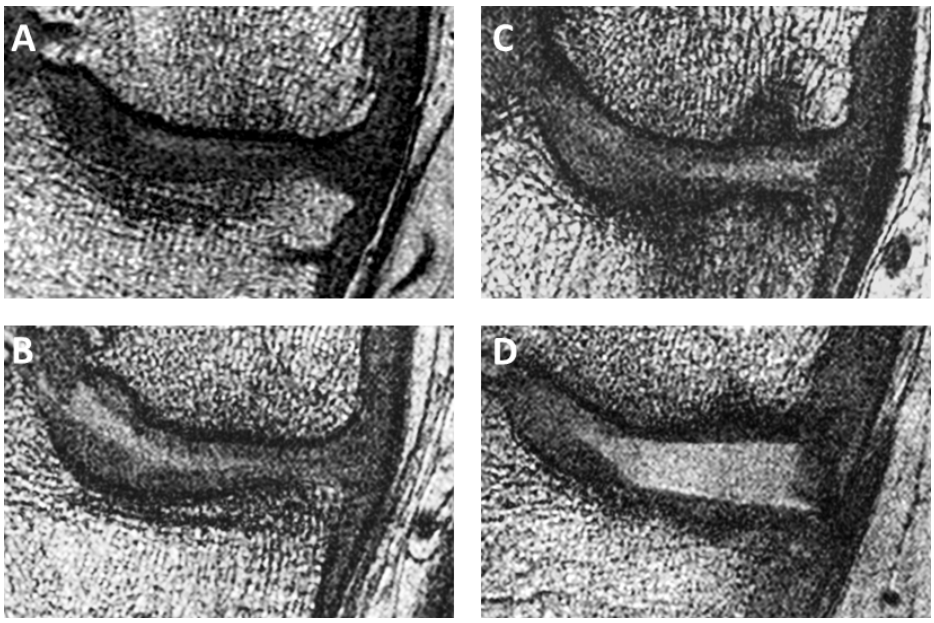
†Odds ratio of being a progressor for each 1 standard deviation increase in texture score.

#Three texture features with largest standardised coefficients (β) in the logistic regression model (note initial tibial model only included two texture features).

Abbreviations: T – tibial feature, F – femoral feature, Gr – Gradient, GrVar – Gradient Variance, ASM – angular second moment.

Figure 4.5

Example coronal 3D FISP MR images through the medial tibiofemoral compartment of radiographic progressors and non-progressors at the initial timepoint. *A*, non-progressor, texture score (TS) -1.22, *B*, non-progressor, TS -1.44, *C*, progressor, TS +1.01, and *D*, progressor, TS +0.93. Higher texture scores correspond to less spatially organized subchondral bone



.

4.4 DISCUSSION

This study demonstrates that combinations of initial and 12-18 month change in subchondral bone MR texture features are significantly associated with radiographic OA progression over 36-months, with better predictive ability for 12-18 month change data.

These results suggest that subchondral bone texture may be a useful quantitative imaging biomarker for use in clinical trials, particularly those with interventions targeting subchondral bone. Although direct interpretation of the texture scores used in this study is difficult, results are consistent with progressors having less spatially organized, more homogeneous subchondral bone at the initial timepoint, and 12-18 month changes in the same direction. While direct visual assessment of what the texture scores are representing is not possible, anecdotally there are cases where visual assessment of example images supports this finding (Figure 4.5). Combinations of texture features were able to predict progressor status using both initial and 12-18 month change in texture features. Classification accuracy improved when tibial and femoral data were combined. However, the best performing model had only modest predictive ability (AUC 0.68), despite the strong associations between texture score and radiographic progression.

Possible explanations for the limited performance of MR subchondral bone texture in this study include a bias towards advanced OA in this cohort, and the MR sequence used. First, as members of the progression subcohort of the OAI, most participants had established OA. The study sample was further biased towards more advanced disease by the fact that only individuals with established OA at the 48 month OAI visit had quantitative JSW measurements performed at the 72-month OAI visit. MR subchondral bone texture has previously demonstrated the ability to detect relatively early OA-related changes in the subchondral bone, therefore it may be that it is of less use in more established disease. Second, the MR sequence used as part of the OAI BAS was optimised for analysing trabecular microarchitecture (i.e. direct estimation of histomorphometry parameters). While texture analysis has previously been described using a similar MR sequence, it has also been demonstrated that alternative MR sequences may provide improved texture discrimination between individuals(224).

Nevertheless, the strength of association between texture scores and radiographic progression and the AUC values presented here are competitive when compared to several alternative OA imaging biomarkers, especially considering the relatively short follow-up time and matching of the control cohort in this study for important predictive covariates of age, sex and BMI(40,52,60,225,226). For example, in the FNIH OA Biomarkers Consortium studies, central medial tibiofemoral compartment cartilage loss over 24 months was associated with radiographic progression at 24-28 months with a similar odds ratio (3.8 [95% CI 2.7 to 5.5]) to that for 12-18 month change in combined texture score (3.8 [2.0 – 7.8]) in this study (60). In the same FNIH cohort, associations between femoral bone shape change (OR 2.7 [2.0 to 3.6]) and femoral bone area change (2.9 [2.1 to 3.9]) over 24 months and radiographic progression at 24-48 months were also of a similar magnitude to those described here(226). Associations between fractal signature analysis (FSA) parameters obtained from plain radiographs and radiographic progression in the FNIH OA Biomarkers cohort were weaker than those described in this study, as was the predictive ability of FSA(40).

The results of this study demonstrate that MR TA of subchondral bone can be considered a useful addition to the suite of imaging biomarkers available for further OA imaging research studies. One advantage over alternative techniques is the multidimensional data output of texture analysis, which is well-placed to interact with machine learning based approaches to image interpretation.

One disadvantage of MR TA is that it is not always clear what the biological or structural correlates of individual texture features are, despite previous demonstration of association with histomorphometry(213). However, as has been shown by the use of texture analysis in other imaging applications, this lack of correspondence to an underlying structural 'ground-truth' does not preclude the use of this method to improve our understanding of the underlying disease process(227–229). In the present study, MR TA has usefully quantified the degree of 'abnormality' in the appearance of the subchondral bone despite limited structural correlation of the texture parameters used. Moreover, based on previous work it is possible to speculate as to the biological meaning of the alterations in subchondral bone texture seen in this study. For example, baseline tibial histogram mean in this study was lower in cases than controls

and was one of the most important features in baseline models for predicting progression. In a previous study, lower values of this parameter were associated with higher bone volume fractions, higher trabecular thickness and lower trabecular spacing suggesting that subchondral bone in progressors in this study was more sclerotic at baseline than that of non-progressors(213).

Future work could evaluate MR TA of subchondral bone in alternative populations, and a head-to-head comparison of different methods for analysing subchondral bone would help to determine the optimal imaging biomarker for use in clinical trials. A barrier to performing this comparison in the present study was the fact that the platforms used for several alternative methods are not freely available, in contrast to the method used here. Automation of the time-consuming ROI drawing procedure would also encourage wider use of this method.

There are several limitations to the present study. Structural correlates for texture parameters have been assessed in previous studies, but using a different MR sequence. Nevertheless, a comparison of texture features between sequences has shown similar changes in both, so it is reasonable to assume similar structural correlates for the texture features derived from the MR sequence used in this study(224). This study used a retrospective case-control design which is subject to selection bias. However, the matching process ensured that cases and controls were well-matched for important baseline characteristics. In common with other longitudinal studies using the OAI dataset, it is not possible to completely separate concurrent from predictive validity for MR subchondral bone texture as the period of follow-up for change in MR texture features overlapped with the follow-up period for radiographic progression. There was a small sample size for the number of texture features analysed which risks introducing bias into any classification procedure. However, this was minimized by excluding texture features with poor reproducibility, using cross-validation, and using penalised regression to limit the number of texture features incorporated into the classification models. Finally, OA progression in this study was defined based on change in radiographic joint space width. This measure is established and robust, but captures only one aspect of OA progression. Symptomatic progression is also important, but not considered in this study due to the low numbers of symptomatic progressors in this cohort.

In conclusion, initial and 12-18 month change in combinations of MR subchondral bone texture features were associated with 36-month radiographic OA progression, with better predictive performance of 12-18 month change data.

APPENDIX 4.1

Inter-observer reproducibility data

Parameter	ICC (95% CI)*	RMSCV (%)
Tibia		
Mean	**0.95 (0.90, 0.98)**	**5.5**
Skewness	0.78 (0.56, 0.90)	24.4
Kurtosis	0.69 (0.40, 0.86)	43.4
Variance	**0.94 (0.87, 0.97)**	**5.5**
Gradient Mean	**0.97 (0.94, 0.99)**	**1.8**
Gradient Skewness	0.90 (0.77, 0.95)	10.9
Gradient Kurtosis	0.63 (0.30, 0.82)	42.4
Gradient Variance	**0.98 (0.96, 0.99)**	**1.4**
Gradient NonZeros	**0.93 (0.83, 0.97)**	**0.9**
Run fraction	**0.95 (0.88, 0.98)**	**0.4**
Short run-length emphasis	**0.94 (0.87, 0.98)**	**0.3**
Long run-length emphasis	**0.93 (0.85, 0.97)**	**1.1**
Grey-level non-uniformity	0.62 (0.29, 0.82)	10.0
Run-length non-uniformity	0.59 (0.25, 0.80)	13.6
Angular Second Moment	**0.88 (0.75, 0.95)**	**8.2**
Contrast	**0.98 (0.95, 0.99)**	**3.9**
Correlation	0.75 (0.47, 0.89)	4.0
Entropy	**0.94 (0.87, 0.98)**	**1.1**
Inverse difference moment	**0.95 (0.88, 0.98)**	**2.2**
Femur		
Mean	**0.95 (0.89, 0.98)**	**3.8**
Skewness	0.88 (0.73, 0.95)	18.8
Kurtosis	0.92 (0.03, 0.97)	43.1
Variance	**0.97 (0.92, 0.99)**	**5.3**
Gradient Mean	**0.98 (0.94, 0.99)**	**1.6**
Gradient Skewness	0.96 (0.92, 0.98)	14.3
Gradient Kurtosis	0.89 (0.77, 0.95)	17.5
Gradient Variance	**0.98 (0.96, 0.99)**	**2.1**
Gradient NonZeros	**0.95 (0.88, 0.98)**	**0.7**
Run fraction	**0.96 (0.91, 0.98)**	**0.3**
Short run-length emphasis	**0.96 (0.91, 0.98)**	**0.2**
Long run-length emphasis	**0.95 (0.89, 0.98)**	**1.0**
Grey-level non-uniformity	0.80 (0.58, 0.91)	9.8
Run-length non-uniformity	0.66 (0.35, 0.84)	11.7
Angular Second Moment	**0.95 (0.88, 0.98)**	**4.2**
Contrast	**0.98 (0.95, 0.99)**	**3.6**
Correlation	0.81 (0.60, 0.91)	8.0
Entropy	**0.96 (0.92, 0.98)**	**0.8**
Inverse difference moment	**0.96 (0.91, 0.98)**	**1.8**

*single measures, absolute agreement
Texture features in **bold** were included in subsequent analyses
Abbreviations: ICC – intraclass correlation coefficient, RMSCV – root mean square average of the coefficients of variation, Gradient NonZeros – proportion of pixels with non-zero gradient

APPENDIX 4.2

Initial and 12-18 month change in individual subchondral bone texture features in case and control participants

Texture feature	Baseline			12-18 month change		
	Cases (mean (SD))	Controls (mean (SD))	OR (95% CI)*	Cases (mean (SD))	Controls (mean (SD))	OR (95% CI)*
Tibia						
Mean	178 (40)	191 (35)	0.77 (0.53, 1.11)	11.8 (42.4)	3.7 (40.5)	1.22 (0.83, 1.87)
Variance	8060 (2890)	9000 (3112)	0.73 (0.50, 1.04)	902 (3198)	501 (2688)	1.15 (0.78, 1.71)
GrMean	1.63 (0.36)	1.68 (0.36)	0.88 (0.61, 1.26)	0.04 (0.49)	0.12 (0.46)	0.84 (0.57, 1.24)
GrVariance	0.95 (0.34)	0.99 (0.35)	0.89 (0.61, 1.27)	0.05 (0.48)	0.10 (0.40)	0.89 (0.60, 1.31)
GrNonZeros	0.88 (0.05)	0.88 (0.06)	0.91 (0.63, 1.30)	0.004 (0.06)	0.02 (0.08)	0.80 (0.53, 1.18)
Run Fraction	0.87 (0.03)	0.87 (0.03)	0.90 (0.63, 1.29)	0.003 (0.04)	0.01 (0.04)	0.81 (0.54, 1.19)
SRLE	0.90 (0.02)	0.90 (0.02)	0.90 (0.62, 1.29)	0.003 (0.03)	0.01 (0.03)	0.81 (0.54 0 1.19)
LRLE	1.52 (0.16)	1.51 (0.18)	1.08 (0.76, 1.56)	-0.01 (0.20)	-0.06 (0.23)	1.24 (0.84, 1.87)
ASM**	3.93 (1.93)	3.72 (2.10)	1.11 (0.78, 1.61)	-0.25 (2.17)	-0.92 (2.68)	1.33 (0.90, 2.03)
Contrast	38.3 (18.0)	40.5 (17.9)	0.88 (0.61, 1.26)	2.3 (25.3)	5.8 (21.6)	0.86 (0.58, 1.26)
Entropy	2.58 (0.20)	2.61 (0.20)	0.88 (0.61, 1.25)	0.03 (0.25)	0.09 (0.25)	0.80 (0.54, 1.18)
Inv Df Mom	0.22 (0.04)	0.21 (0.04)	1.13 (0.79, 1.63)	-0.005 (0.05)	-0.02 (0.06)	1.25 (0.85, 1.86)
Femur						
Mean	176 (34)	187 (33)	0.70 (0.47, 1.00)	13.8 (36.5)	8.3 (28.4)	1.18 (0.80, 1.78)
Variance	6940 (2677)	7433 (2122)	0.87 (0.60, 1.24)	1086 (2440)	551 (2221)	1.27 (0.86, 1.93)
GrMean	1.67 (0.36)	1.72 (0.38)	0.89 (0.62, 1.27)	0.07 (0.45)	0.17 (0.48)	0.80 (0.53, 1.18)
GrVariance	0.97 (0.35)	1.00 (0.36)	0.92 (0.64, 1.31)	0.06 (0.48)	0.14 (0.45)	0.82 (0.55, 1.22)
GrNonZeros	0.89 (0.05)	0.89 (0.06)	0.94 (0.65, 1.34)	0.009 (0.05)	0.03 (0.07)	0.78 (0.51, 1.15)
Run Fraction	0.87 (0.03)	0.88 (0.03)	0.92 (0.64, 1.31)	0.006 (0.03)	0.003 (0.04)	0.78 (0.52, 1.15)
SRLE	0.90 (0.02)	0.91 (0.03)	0.92 (0.64, 1.31)	0.004 (0.02)	0.003 (0.03)	0.77 (0.51, 1.14)
LRLE	1.51 (0.15)	1.50 (0.18)	1.06 (0.74, 1.53)	-0.03 (0.16)	-0.08 (0.23)	1.29 (0.87, 1.96)
ASM**	4.27 (2.35)	4.27 (2.70)	1.00 (0.70, 1.44)	-0.45 (2.24)	-1.31 (3.34)	1.37 (0.92, 2.12)
Contrast	38.6 (17.5)	40.5 (18.6)	0.90 (0.62, 1.29)	3.0 (23.8)	8.1 (22.8)	0.80 (0.53, 1.18)
Entropy	2.55 (0.20)	2.56 (0.22)	0.93 (0.65, 1.34)	0.05 (0.24)	0.11 (0.28)	0.78 (0.52, 1.15)
Inv Df Mom	0.21 (0.04)	0.21 (0.05)	1.10 (0.77, 1.59)	-0.009 (0.05)	-0.005 (0.05)	1.30 (0.88, 1.97)

*odds ratio (OR) and 95% confidence interval (95% CI) of being a case for a 1 standard deviation increase in texture feature

**values as given x10-3

Abbreviations: Gr – Gradient, GrNonZeros – proportion of pixels with non-zero gradient, SRLE – short run-length emphasis, LRLE – long run length emphasis, ASM – angular second moment, Inv Df Mom – inverse difference moment.

CHAPTER 5:

THE ADVANCED MAGNETIC RESONANCE IMAGING OF OSTEOARTHRITIS (AMROA) STUDY – CORE METHODOLOGY & RESULTS

The AMROA study was a prospective longitudinal observational study conducted with the aim of assessing the potential utility of quantitative imaging biomarkers (QIBs) of bone, cartilage and synovium in the setting of experimental medicine studies. Candidate QIBs were compared across the domains of test-retest repeatability, ability to discriminate between participants with osteoarthritis (OA) and age-matched healthy controls, and responsiveness to change over 6 months and 1 year.

This chapter describes the core background, methodology and results of the study, including the results of radiographic and semi-quantitative MR analyses. Subsequent chapters will discuss the different QIB methods and results in detail.

5.1 INTRODUCTION

As discussed in previous chapters, multiple quantitative imaging biomarkers (QIBs) of OA have been described which have the potential to be of use in early phase clinical studies. However, the implementation of QIBs in this type of study remains limited. One of the possible reasons for their limited uptake is a relative paucity of data on the performance characteristics of candidate QIBs in this setting.

This chapter describes the Advanced Magnetic Resonance imaging of Osteoarthritis (AMROA) study which was designed to help address this shortcoming. AMROA was a prospective longitudinal observational study aiming to evaluate the performance characteristics of a selection of QIBs in conditions similar to an early phase 'experimental medicine' study. Experimental medicine studies are typically biomarker-rich studies occurring early in the drug development process(76). They aim to establish early proof-of-concept or proof-of-mechanism, thus de-risking subsequent development and contributing to the decision of whether to progress a novel therapeutic entity into later stage development (so-called 'go/no-go' decision making) with the entailed resource implications(75). Experimental medicine studies overlap with phases 1 and 2a of the traditional drug development paradigm and are typically of short duration with small numbers of participants. This poses particular challenges for imaging biomarkers in OA, traditionally considered to be a slowly progressive disorder.

The AMROA study included candidate QIBs which both reflected the multi-tissue nature of OA and had also demonstrated potential in larger studies (Table 5.1). For synovitis QIBs, evidence of potential principally comes from studies evaluating response to intra-articular steroids using similar methodology(71,72,230). For cartilage QIBs, evidence of promise of the approach outlined in chapter 3 comes from previous work in larger observational datasets(146,187,188). Bone QIBs similar to those employed in this study have been shown to be useful in the Osteoarthritis Initiative dataset, including the analysis described in Chapter 4(64,226).

Table 5.1

Description of QIBs included in AMROA study

Tissue	Method	QIBs
Synovium	DCE-MRI	K^{trans}
		$IAUC_{60}$
	CE-MRI	Synovial tissue volume
Cartilage	3D-CaSM (chapter 3)	Cartilage thickness
		T1rho relaxation
		T2 relaxation
		dGEMRIC
Bone	3D-CaSM (chapter 3)	Bone area*
	Texture analysis (chapter 4)	Bone texture

Abbreviations: DCE-MRI – dynamic contrast-enhanced MR imaging, CE-MRI – contrast-enhanced MR imaging, 3D-CaSM – 3D-cartilage surface mapping, dGEMRIC – delayed gadolinium enhanced MR imaging of cartilage.
*tAB including osteophytes per Eckstein nomenclature(231)

The purpose of the AMROA study was to evaluate the likely performance of candidate QIBs in the setting of experimental medicine studies. Specifically, I aimed to assess QIBs across the domains of test-retest repeatability, ability to discriminate between OA and healthy controls, and responsiveness to change over 6-months. This chapter outlines the core (non-QIB) methodology and results for the study. Subsequent chapters discuss the methods and results for specific QIBs of each joint tissue.

5.2 METHODS

5.2.1 Ethical approval

Ethical approval for the study was granted by the Cambridge Central Local Research Ethics Committee (ref: 16/EE/0402). All participants gave written, informed consent.

5.2.2. Participants

Participants with knee OA and healthy volunteers (HV) were recruited. OA participants were recruited from specialist orthopaedic knee clinics at Addenbrooke's Hospital. HVs were recruited either via a register of healthy individuals who had expressed an interest in taking part in clinical studies, or via poster advertisements placed in the University of Cambridge Department of Radiology.

Inclusion criteria for both groups were age 40-60 years, in good general health, body mass index ≤ 35 kg/m^2 and the imaged knee able to fit within the knee MR coil (diameter approximately 18 cm). Additional inclusion criteria for OA participants were a clinical diagnosis of OA per American College of Rheumatology criteria as assessed by a consultant orthopaedic surgeon and medial compartment predominant disease with a Kellgren-Lawrence grade of 2-3 as assessed on standing AP knee radiographs by an orthopaedic surgeon and musculoskeletal radiologist in consensus(27,232). Exclusion criteria for both groups were a history of previous ipsilateral lower limb fracture or surgery (including arthroscopy), history of metabolic bone disease or inflammatory arthritis or contraindication to MR imaging or gadolinium-based contrast agent (gdCA) administration.

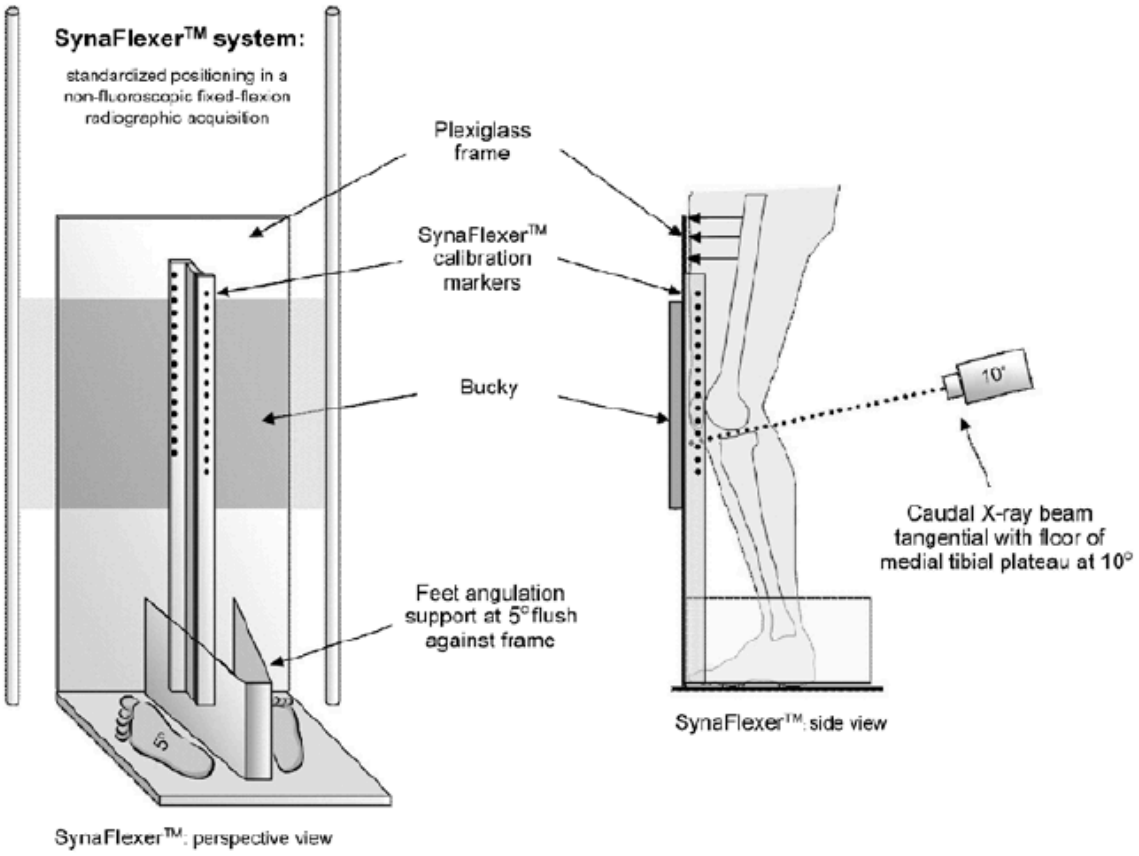
5.2.3 Study Procedures

Participants had four scheduled study visits: baseline, 1-month, 6-months and 1-year. Each participant had an MR examination according to the protocol described below at each visit. OA participants also had radiographs performed at baseline and at 1-year. These were performed posteroanterior (PA) in fixed flexion with the use of a SynaFlexer (BioClinica, Newtown, PA) positioning device to ensure a consistent degree of knee flexion between examinations(233). For baseline radiographs, the x-ray beam was initially angled 10^o caudally (Figure 5.1). Images were immediately reviewed to ensure adequate superimposition of the anterior and posterior margins of the tibial

plateau, defined as less than 0.5 mm difference between the two. If superimposition was unsatisfactory, the radiograph was repeated with a change in the caudal angulation of the x-ray beam in 2° increments in the direction of the perceived discrepancy. Up to four radiographic exposures were permitted at each visit. Each participant had their baseline x-ray beam angulation recorded which was used as the initial angulation for the 1-year follow-up radiograph.

Figure 5.1

Fixed flexion radiographic technique using SynaFlexer frame. Reproduced from (233).



The Knee Injury and Osteoarthritis Outcome Score (KOOS) questionnaire was completed by OA participants at each visit(234). HV participants completed the KOOS questionnaire at the baseline visit only. The KOOS questionnaire used for the study is provided in Appendix 5.1. At each visit, a whole blood sample was taken from participants to enable measurement of haematocrit for DCE-MR analyses. Serum samples were also taken and frozen to enable future measurement of exploratory serum biomarkers (e.g. cartilage turnover markers) and assess their correlation to imaging findings.

MR examinations were scheduled consistently at the same time of day for each visit of each participant to minimise the confounding effect of diurnal changes in cartilage thickness and composition measurements(128,235). In addition, all participants had a period of unloading of the knee of at least 45 minutes prior to imaging to minimise effects on the same measures related to recent weight-bearing(236,237).

All MR examinations were performed on a 3T platform (GE 750; GE Healthcare, Waukesha, WI) using an 8-channel transmit/receive knee coil (Invivo, Gainesville, FL). Careful immobilisation of the target knee was performed to minimise motion artefact. OA participants had the symptomatic knee imaged. HV participants had a knee selected at random using a random number generator (www.random.org) by the supervising radiologist immediately prior to the baseline visit.

The MR protocol was divided into two sessions, the first lasting approximately 45 minutes and the second lasting approximately five minutes (Table 5.2). Towards the end of the first session, the DCE-MR sequences were performed. GdCA (Dotarem®, Guerbet, Paris, France) was administered via a pump injector at a rate of 3 mL/s followed by 50 mL saline administered at the same rate. The dose of gdCA was 0.2 mM/kg, double the normal dose for the purposes of subsequent dGEMRIC. The additional steps for dGEMRIC included participants leaving the scanner after the first MR session and performing 10 minutes of cycling at an easy pace on a stationary cycle in a room adjacent to the MR control room. There was then a rest period of approximately 70 minutes, followed by the second MR session.

Table 5.2

AMROA study MR protocol

Purpose	Sequence	TR/TE (ms)	Flip angle(°)	Acquired matrix (pixels)	FOV (mm)	Slice thickness/gap (mm)	NEX	Acquisition time (min:s)	Notes
SESSION 1									
Standard clinical assessment & MOAKS scoring	Sag IW FS FSE	1500/36	90	384 x 256**	160	3/1	3	03:00	
	Cor IW FS FSE	2500/36	90	384 x 256**	160	3/1	3	03:30	
	Cor T1 FSE	600/12	90	512 x 512	120	2.5/2.5	1	02:30	Used for subchondral bone texture analysis
Cartilage morphology	3D SPGR FS	26/in-phase	25	512 x 380**	150	1/0	0.5	06:50	
Cartilage composition	3D FSE FS CubeQuant T1ρ	1500/1,10,20,35*	90	320 x 256**	160	3/0	0.5	05:00	Spin lock frequency 500 Hz
	3D FSE FS CubeQuant T2 3D FSE FS	1500/7,13,27,40	90	320 x 256**	160	3/0	0.5	05:00	
Synovial segmentation	3D SPGR FS	12/in-phase	15	384 x 384**	320	2[#]/0	0.5	04:30	
T1 map for DCE-MRI	3D SPGR	4/in-phase	2,6,14	128 x 128[†]	320	4[‡]/0	0.5	04:00	6 phases per flip angle
IV Dotarem administered during next sequence after first 5 phases									
DCE-MRI	3D SPGR	4/in-phase	14	128 x 128[†]	320	4[‡]/0	0.5	08:10	35 phases, 14 s/phase
Synovial segmentation	3D SPGR FS	12/in-phase	15	384 x 384**	320	2[#]/0	0.5	04:30	
SESSION 2									
Cartilage composition (dGEMRIC)	3D SPGR	5/in-phase	2,6,14	320 x 256**	160	3/0	1	03:30	

Abbreviations: TR – repetition time, TE – echo time, FOV – field-of-view, NEX – number of excitations (signal averages), IW – intermediate-weighted, FS – fat saturated (spectral), FSE – fast spin echo, SPGR – spoiled gradient echo, IV – intravenous

* Time of spin lock (TSL) values.

** Interpolated to 512 x 512 with zero-filling [†] Interpolated to 256 x 256 with zero-filling [#] Interpolated to 1 mm with zero-filling [‡] Interpolated to 2 mm with zero-filling

5.2.4 Non-imaging assessments

Demographic data were acquired for each participant at baseline including age, sex and body mass index (BMI). KOOS scores were recorded for each individual subscale (pain, symptoms, activities of daily living, sports & recreation and knee-related quality of life) as well as the average score across all five subscales, KOOS-5.

Whole blood samples were processed by the Cambridge University Hospitals NHS Foundation Trust laboratories. Haematocrit values were recorded and used for subsequent DCE-MR analyses(238). Serum blood samples were centrifuged with subsequent aliquoting and freezing of the serum fraction at -80° Celsius.

5.2.5 Core imaging assessments

Fixed-flexion radiographs were graded according to the Kellgren-Lawrence scale, with additional grading of joint space narrowing and osteophytes according to the Osteoarthritis Research Society International (OARSI) scale(27,28). Fixed location joint space width was measured manually at 22.5% of the total width of the tibiofemoral joint (JSW_{225}), defined by lines tangent to the most medial/lateral point on the femoral condyles on the PA radiograph (Figure 1.2). Measurement at this position has been demonstrated to be the most repeatable and sensitive to change when using the fixed-flexion technique(33).

The standard clinical sequences from the MR protocol (sagittal intermediate weighted (IW) fat-saturated (FS), coronal IW FS and coronal T1) plus the high spatial resolution 3D spoiled gradient echo (SPGR) sequence were used to grade participants according to the MRI osteoarthritis knee score (MOAKS) system(46). Whether semi-quantitative reading should be blinded to timepoint is a matter of ongoing debate. Blinding to timepoint reduces bias but also reduces sensitivity to change(239,240). The general consensus in OA imaging is that not performing blinding to timepoint is probably preferable, therefore blinding was not performed in this study (241). Quantitative MR assessments will be discussed in detail in subsequent chapters, but, in brief, these included measurement of K^{trans}, $IAUC_{60}$ and synovial tissue volume, cartilage thickness and composition (T1rho, T2 and dGEMRIC) and subchondral bone area and texture.

5.2.6 Statistical analysis

Demographic, clinical (KOOS), radiographic and semiquantitative MR imaging data were analysed descriptively, using means (standard deviations) or medians (interquartile ranges) as appropriate. Formal hypothesis testing was not performed (a) as these variables were not the primary outcome measures of interest and (b) due to the exploratory nature of the study.

For KOOS scores, test-retest repeatability was assessed by calculating the root-mean-square average coefficient of variation (RMSCV) between baseline and 1-month measurements in OA participants. The proportion of OA participants with significant improvement or worsening of symptoms over 6-months and 1-year was calculated using thresholds of the smallest detectable difference (SDD, 20 points) and also OMERACT-OARSI responder criteria (242,243).

The proportion of OA participants with radiographic progression over 1-year was calculated. For qualitative grading scales (Kellgren-Lawrence and OARSI), progression was defined as any increase in grade on the corresponding scale. For JSW_{225} measurement, progression was defined as a reduction in JSW of ≥ 0.7 mm based on criteria developed by the Foundation of the National Institutes of Health (FNIH) OA biomarkers consortium as having a less than 10% chance of being due to measurement error(226).

Participants were assessed for improvements or worsening in their MOAKS score over 6-months and 1-year using previously published criteria(244). This assessment was performed separately for each tissue type scored by MOAKS for which a change definition was available (bone marrow lesions, cartilage defects, synovitis, osteophytes meniscal pathology and meniscal extrusion).

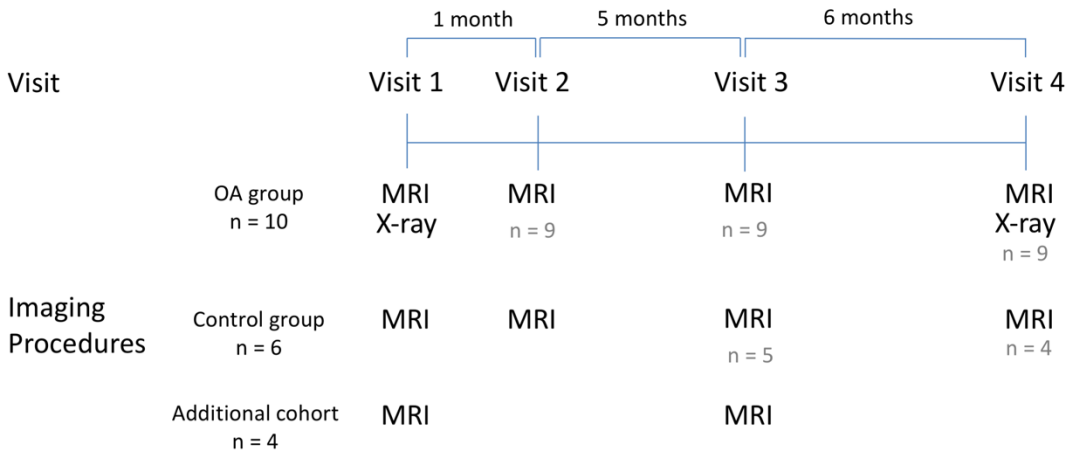
5.3 RESULTS

5.3.1 Participants

An original cohort of 10 OA participants and six HVs were recruited. All original cohort participants completed the baseline study visit, nine OA/six HV participants completed the 1-month study visit, nine/five completed the 6-month study visit, and nine/four completed the 1-year study visit. An informal review by the study team of the 1-month test-retest repeatability data recommended the recruitment of up to five additional participants with OA to be imaged at baseline and six months. This was due to the smallest detectable difference (SDD) values for QIBs calculated from these data being smaller than anticipated, increasing the likelihood of detecting meaningful changes at six months. Additional participants were therefore sought to increase the confidence in estimates of magnitude and standard deviation of change at this timepoint. Consequently, an additional cohort of four OA participants were recruited and imaged at baseline and six months (Figure 5.2). All additional cohort participants completed both study visits.

Figure 5.2

Study visit schedule. All participants in each group completed each visit unless otherwise stated in grey.



Baseline demographic data are provided in Table 5.3. Nine OA participants underwent arthroscopy during the study follow-up period (eight in the original cohort, one in the additional cohort). No participants were treated with any investigational therapeutic agents, and none underwent administration of any intra-articular medication.

Table 5.3

Baseline demographic characteristics of included study participants

	OA (n = 14)	**HV (n = 6)**
Age (years)*	51(5)	55(4)
Sex (M:F)	8:6	2:4
BMI(kg/m^2)*	29.2 (4.1)	29.7 (3.0)

*mean (SD)

5.3.2 Non-imaging assessments

KOOS scores for OA and HV participants are provided in Table 5.4. All HV participants had KOOS scores above the population mean for their age and sex strata, supporting their use as control participants in the study(245).

Table 5.4

KOOS scores for OA and HV participants. Values provided are means (SDs).

	OA				HV
	Baseline (n = 14)	1-month (n = 9)	6-month (n = 13)	1-year (n = 8)	Baseline (n = 6)
Pain	60 (20)	56 (19)	74 (15)	64 (19)	99 (2)
Symptoms	56 (19)	62 (22)	77 (18)	67 (19)	96 (5)
ADL	67 (21)	60 (22)	82 (16)	71 (18)	99 (1)
Sports/Rec	36 (23)	37 (26)	56 (24)	41 (27)	95 (6)
QoL	33 (23)	28 (20)	50 (20)	35 (22)	96 (8)
KOOS-5	50 (18)	48 (19)	68 (15)	56 (19)	97 (4)

Abbreviations: ADL – activities of daily living, QoL – knee-related quality of life, Sports/Rec – sports & recreation, KOOS-5 – average KOOS score across all 5 subscales.

One-month test-retest RMSCV for the KOOS subscales of pain, symptoms, activities of daily living, sports and recreation and quality of life were 10%, 18%, 6%, 35% and 25% with a RMSCV for the KOOS-5 average score of 12%. These are commensurate with previously published values for the KOOS(242).

The number of improvers based on the SDD and OMERACT-OARSI responder thresholds are provided in Table 5.5. No participants demonstrated worsening of KOOS values greater than the SDD.

Table 5.5

Number (%) of OA participants with symptomatic improvement at 6-months and 1-year visits per SDD and OMERACT-OARSI criteria.

	6-month follow-up (n = 14)		1-year follow-up (n = 8)	
Change criterion	SDD	O-O	SDD	O-O
Pain	4 (29%)	5 (36%)	3 (38%)	1 (13%)
Symptoms	6 (43%)		2 (25%)	
ADL	5 (36%)		3 (38%)	
Sports/Rec	6 (43%)		4 (50%)	
QoL	7 (50%)		4 (50%)	
KOOS-5	5 (36%)		3 (38%)	

Values provided are participant numbers. Note that the O-O criterion uses a composite of different measurement scales, more details in (243). **Abbreviations:** SDD – smallest detectable difference, O-O – OMERACT-OARSI

5.3.4 Core Imaging Assessments

All of the original OA cohort participants had baseline radiographs, with seven out of 10 also completing 1-year follow-up radiographs. One participant did not attend their final visit, with the other two participants missing their final radiograph due to scheduling difficulties. Individual participant qualitative and quantitative radiographic assessments are provided in Table 5.6 for both original and additional cohort participants. Additional cohort participants did not have dedicated fixed-flexion radiographs making measurement of JSW unreliable, therefore only qualitative assessments were performed.

Table 5.6

Results of radiographic analyses in OA participants

ID	K-L grade		OARSI joint space narrowing		OARSI osteophytes		JSW_{225} (mm)	
	Baseline	1-year	Baseline	1-year	Baseline	1-year	Baseline	1-year
Original cohort								
1	2	2	1	1	1	1	5.2	4.7
2	3	3	2	2	1	1	3.5	2.5*
3	2	2	1	1	1	1	4.5	4.1
4	2	-	2	-	1	-	3.9	-
5	3	3	3	3	2	2	3.3	3.3
6	2	2	1	1	1	1	5.4	4.8
7	3	3	2	2	1	1	4.0	3.8
8	3	3	3	3	2	2	3.1	3.6
9	2	-	0	-	1	-	6	-
10	2	-	0	-	1	-	6.3	-
Additional cohort								
11	2	-	0	-	1	-	-	-
12	2	-	1	-	1	-	-	-
13	2	-	1	-	1	-	-	-
14	2	-	1	-	2	-	-	-

* JSW progression per criteria in section 5.2.9

One out of seven analysable original cohort participants had 1-year JSW loss greater than the specified 0.7 mm threshold. There was no change in K-L or OARSI grades over 1-year in analysable original cohort participants.

Semi-quantitative assessment using MOAKS was possible for each participant at each completed study visit. Baseline MOAKS data are summarised in Table 5.7. MOAKS progression data are presented in Figure 5.3.

Table 5.7

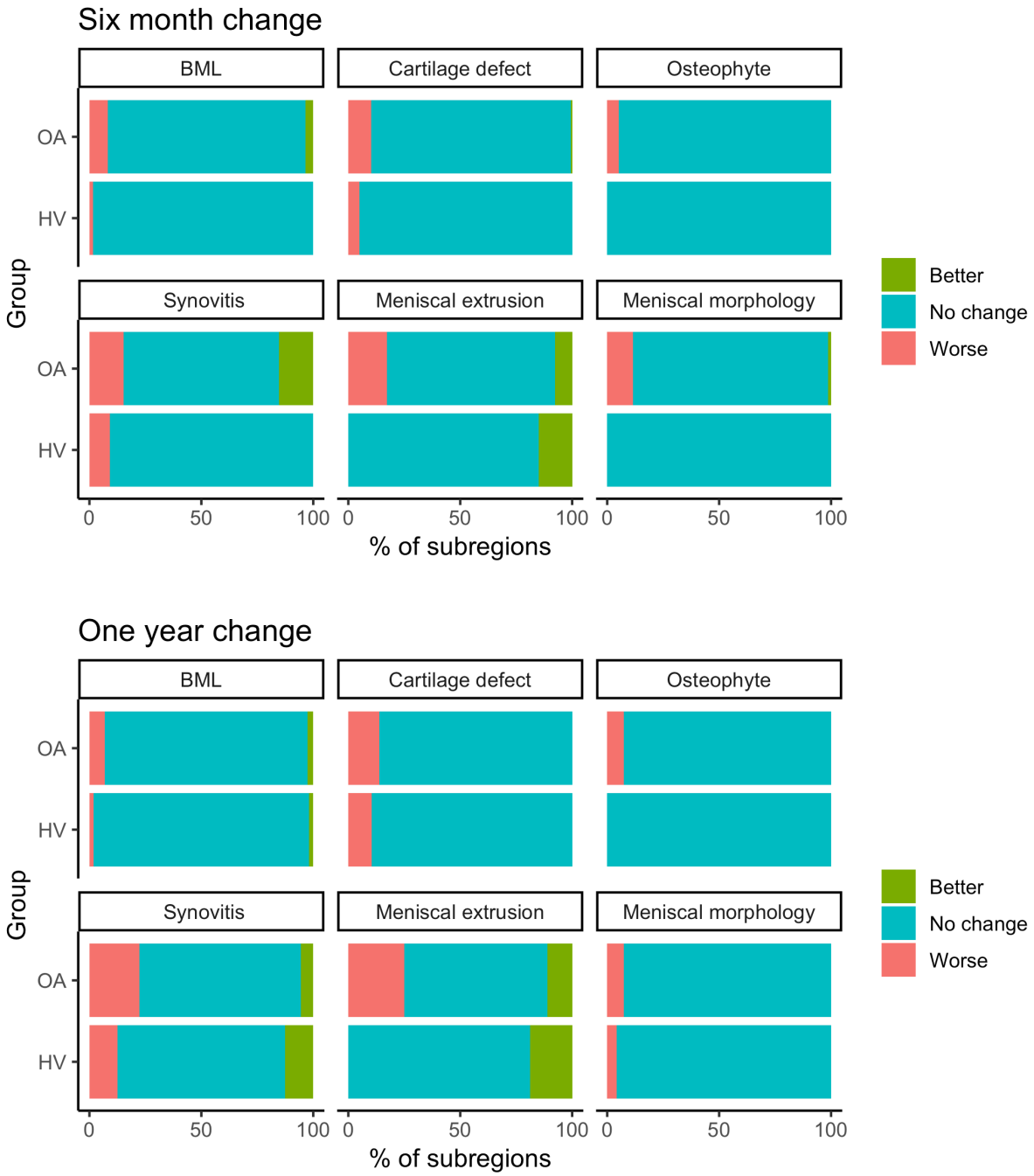
Summary of baseline MOAKS data

MOAKS domain	OA (n = 14)	HV (n = 6)
Bone marrow lesion		
- Any	12	1
- ≥ grade 2	8	0
Cartilage defect		
- Any	12	2
- ≥ grade 2 (size)	12	1
- Any full-thickness	9	1
Osteophyte		
- Any	12	2
- ≥ grade 2	5	0
Meniscal damage		
- Any medial tear*	10	1
- Any lateral tear*	8	1
Meniscal extrusion		
- Any medial	13	4
- ≥ grade 2 medial	12	4
- Any lateral	1	1
- ≥ grade 2 lateral	1	0
Effusion-synovitis		
- Any	11	1
- ≥ grade 2	5	0
Hoffa-synovitis		
- Any	6	1
- ≥ grade 2	1	0

* not including intrameniscal signal without discrete tear

Figure 5.3

Percentage of subregions in each group with worsening, improvement or no change in MOAKS score over 6-months and 1-year, divided by MOAKS domain. Note synovitis domain is a composite of effusion-synovitis and Hoffa-synovitis



5.4 DISCUSSION

This chapter has described the core methodology and results of non-QIB analyses in the AMROA study. Meaningful changes in semi-quantitative MR grading scores, radiographic JSW measurement and symptoms were observed during the study period.

Participants with OA demonstrated a greater number of MOAKS abnormalities at baseline and an increased proportion of subregions demonstrating progression at 6-months and 1-year when compared to HV participants. Given that one would expect QIBs to be at least as sensitive to change as semi-quantitative or qualitative measurements, this finding suggests that QIBs will also demonstrate meaningful changes in some individuals over the study period. Only one participant demonstrated meaningful radiographic changes over one year, in keeping with previous studies over similar time periods and highlighting the limited utility of JSW measurement in the experimental medicine setting(246).

Not all HV participants had structurally normal knees at MR imaging. Abnormalities such as those detected are common in the asymptomatic population and become increasingly common with age(42). The AMROA study was designed to match HV participants as closely as possible to OA participants for age given that one of the desirable criteria for a candidate QIBs in OA is that it should be able to discriminate between true disease and normal ageing(247).

The number of OA participants demonstrating symptomatic improvement over the study period is likely a result of several factors. The natural waxing and waning nature of symptoms in OA is well-described, although if this was the sole explanation one might expect to see individuals with worsening of symptoms as well. Regression to the mean may also play a role. The local pathway for referral to an orthopaedic knee clinic requires the patient to have persistent, severe symptoms which have not responded to conventional treatments. Therefore, patients are likely to have been referred to orthopaedic clinic at a time when their symptoms were severe, with regression to the mean making an improvement in symptoms likely at subsequent visits. No treatment with known disease-modifying benefit was administered to participants during the study, however symptomatic benefit from treatments such as arthroscopy and simple analgesia are feasible. The improvement in symptoms demonstrated in this study in

the absence of any disease-modifying intervention highlights the difficulties faced when designing trials comparing active treatment vs placebo using patient reported outcome measures alone.

The proportion of participants demonstrating MOAKS progression or improvement is similar to what has been observed in previous studies using the same definitions of change(248,249). Synovitis and meniscal extrusion measures had the greatest proportion of subregions demonstrating change. Synovitis is known to fluctuate over time in knee OA(250,251). Fluctuation in meniscal extrusion is less well-described. However, extrusion of the meniscus is known both to occur in both healthy and abnormal knees and to be influenced by multiple factors including anatomic variation in the site of meniscal root attachment, recent loading and volume of joint effusion, providing a rationale for why fluctuation in measures of meniscal extrusion may have occurred in this study(252–254).

The fact that 9 out of 14 OA participants underwent knee arthroscopy during the study follow-up period is somewhat surprising, given that this procedure is no longer recommended in the setting of knee OA(8). However, this recommendation remains contested with the suggestion that the evidence it is based on is not generalisable to the typical population seen in orthopaedic clinics(9,10). Recently published British Association for Surgery of the Knee (BASK) guidelines on the appropriateness of arthroscopic meniscal surgery in the presence of OA would support the use of this treatment in the patient group enrolled in this study – symptomatic mild-to-moderate OA who have not responded to conservative treatment(11). This uncertainty may explain the fact that a large proportion of participants underwent this procedure and highlights the paucity of effective treatments on offer for the study demographic(12).

There are several strengths of the AMROA study design. A standardised protocol was used to minimise common measurement confounds, including the use of a positioning device for knee radiographs, scheduling MR examinations at a consistent time of day and allowing for a period of unloading prior to compositional knee MR sequences. In addition, all studies were performed at a single centre under direct radiologist supervision. Another strength is the use of an age-matched control cohort. Chapter 2 of this thesis illustrates the fact that the discriminative performance of imaging biomarkers may be inflated when non-age-matched control cohorts are used.

There are several study limitations. Only a small number of participants were included with a short follow-up period. This was intentional in order to mimic the conditions of an experimental medicine study; however, this does limit study power. Due to this constraint, study statistical analyses are predominantly descriptive rather than inferential. While the study had an observational design, several participants had surgical interventions as part of ongoing care during the study period so the responsiveness of different QIBs observed in the study may be different to what would be observed in an interventional trial population. This limitation was unavoidable due to ethical concerns regarding asking participants to forego treatment for a purely observational study without direct benefit. Finally, the careful standardisation of study procedures previously described as a strength is also a potential weakness as it limits the extrapolation of results to less controlled settings such as multi-centre studies.

In conclusion, this chapter has described the core methodology and results of the AMROA study. Future chapters will discuss the methodology and results of QIBs employed in this study in detail.

APPENDIX 5.1

KOOS questionnaire

KOOS KNEE SURVEY

Today's date: _____/_____/_____ Date of birth: _____/_____/_____

Name: _____

INSTRUCTIONS: This survey asks for your view about your knee. This information will help us keep track of how you feel about your knee and how well you are able to perform your usual activities.
Answer every question by ticking the appropriate box, only one box for each question. If you are unsure about how to answer a question, please give the best answer you can.

Symptoms
These questions should be answered thinking of your knee symptoms during the **last week**.

S1. Do you have swelling in your knee?

Never	Rarely	Sometimes	Often	Always
☐	☐	☐	☐	☐

S2. Do you feel grinding, hear clicking or any other type of noise when your knee moves?

Never	Rarely	Sometimes	Often	Always
☐	☐	☐	☐	☐

S3. Does your knee catch or hang up when moving?

Never	Rarely	Sometimes	Often	Always
☐	☐	☐	☐	☐

S4. Can you straighten your knee fully?

Always	Often	Sometimes	Rarely	Never
☐	☐	☐	☐	☐

S5. Can you bend your knee fully?

Always	Often	Sometimes	Rarely	Never
☐	☐	☐	☐	☐

Stiffness
The following questions concern the amount of joint stiffness you have experienced during the **last week** in your knee. Stiffness is a sensation of restriction or slowness in the ease with which you move your knee joint.

S6. How severe is your knee joint stiffness after first wakening in the morning?

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

S7. How severe is your knee stiffness after sitting, lying or resting **later in the day**?

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

Pain

P1. How often do you experience knee pain?

Never	Monthly	Weekly	Daily	Always
☐	☐	☐	☐	☐

What amount of knee pain have you experienced the **last week** during the following activities?

P2. Twisting/pivoting on your knee

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

P3. Straightening knee fully

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

P4. Bending knee fully

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

P5. Walking on flat surface

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

P6. Going up or down stairs

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

P7. At night while in bed

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

P8. Sitting or lying

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

P9. Standing upright

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

Function, daily living

The following questions concern your physical function. By this we mean your ability to move around and to look after yourself. For each of the following activities please indicate the degree of difficulty you have experienced in the **last week** due to your knee.

A1. Descending stairs

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

A2. Ascending stairs

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

For each of the following activities please indicate the degree of difficulty you have experienced in the **last week** due to your knee.

A3. Rising from sitting

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

A4. Standing

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

A5. Bending to floor/pick up an object

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

A6. Walking on flat surface

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

A7. Getting in/out of car

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

A8. Going shopping

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

A9. Putting on socks/stockings

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

A10. Rising from bed

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

A11. Taking off socks/stockings

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

A12. Lying in bed (turning over, maintaining knee position)

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

A13. Getting in/out of bath

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

A14. Sitting

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

A15. Getting on/off toilet

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

For each of the following activities please indicate the degree of difficulty you have experienced in the **last week** due to your knee.

A16. Heavy domestic duties (moving heavy boxes, scrubbing floors, etc)

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

A17. Light domestic duties (cooking, dusting, etc)

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

Function, sports and recreational activities

The following questions concern your physical function when being active on a higher level. The questions should be answered thinking of what degree of difficulty you have experienced during the **last week** due to your knee.

SP1. Squatting

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

SP2. Running

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

SP3. Jumping

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

SP4. Twisting/pivoting on your injured knee

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

SP5. Kneeling

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

Quality of Life

Q1. How often are you aware of your knee problem?

Never	Monthly	Weekly	Daily	Constantly
☐	☐	☐	☐	☐

Q2. Have you modified your life style to avoid potentially damaging activities to your knee?

Not at all	Mildly	Moderately	Severely	Totally
☐	☐	☐	☐	☐

Q3. How much are you troubled with lack of confidence in your knee?

Not at all	Mildly	Moderately	Severely	Extremely
☐	☐	☐	☐	☐

Q4. In general, how much difficulty do you have with your knee?

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

Thank you very much for completing all the questions in this questionnaire.

CHAPTER 6:

DYNAMIC CONTRAST ENHANCED MAGNETIC RESONANCE IMAGING OF SYNOVITIS IN THE AMROA STUDY

Synovial inflammation is common in knee OA. Dynamic contrast enhanced (DCE) MR imaging provides a method of quantifying synovitis. DCE-MR is well established in rheumatoid arthritis and has shown initial promise in OA.

In this chapter I describe the use of DCE-MR quantitative imaging biomarkers (QIBs) of synovitis in the AMROA study. No prior description of test-retest repeatability of DCE-MR QIBs in knee OA or comparison of DCE-MR QIB values between OA participants and age-matched healthy controls exists in the published literature. The results suggest that the DCE-MR QIB with the best performance characteristics for experimental medicine studies is K^{trans}, the transfer coefficient between plasma and the extracellular extravascular space.

6.1 INTRODUCTION

Synovial inflammation is common in OA, with MR-detected synovitis occurring in up to 90% of OA knees(255). It can be detected both histologically and on imaging from the early stages of the disease process(256). Strong cross-sectional associations exist between the presence of synovitis and the severity of knee pain(257). Longitudinal associations have been demonstrated between the presence and severity of synovitis and OA progression from both a symptomatic and a structural point of view(250,258,259). There is therefore a strong rationale for therapeutic targeting of synovitis to provide disease modification(24).

Dynamic contrast enhanced (DCE) MR imaging aims to characterise the uptake and washout of gadolinium-based contrast agents (gdCA) in tissues of interest, providing measures of tissue perfusion, capillary permeability and blood and interstitial volume. These parameters are known to change in the synovium in OA(260). DCE-MR has been used to assess synovitis in early phase clinical trials of rheumatoid arthritis (RhA) and has demonstrated superiority over semi-quantitative assessments in this setting(261,262). The promise of DCE-MR in OA has been illustrated by several studies demonstrating the response of DCE-MR biomarkers to intra-articular corticosteroid treatment with improved responsiveness compared to alternative semi-quantitative and qualitative assessments of synovitis(71,72).

However, to increase confidence in the utility of DCE-MR QIBs in experimental medicine (EM) studies of knee OA more information is desirable. This includes assessment of test-retest repeatability, ability to discriminate between knee OA and normal ageing, and expected changes over relevant follow-up periods.

Therefore, the purpose of this study was to assess the performance characteristics of DCE-MR QIBs of synovitis for use in EM studies.

6.2 METHODS

6.2.1 Participants & Ethical Approval

Ethical approval, participant recruitment and eligibility criteria are provided in chapter 5 (sections 5.2.1 and 5.2.2).

6.2.2 Image acquisition

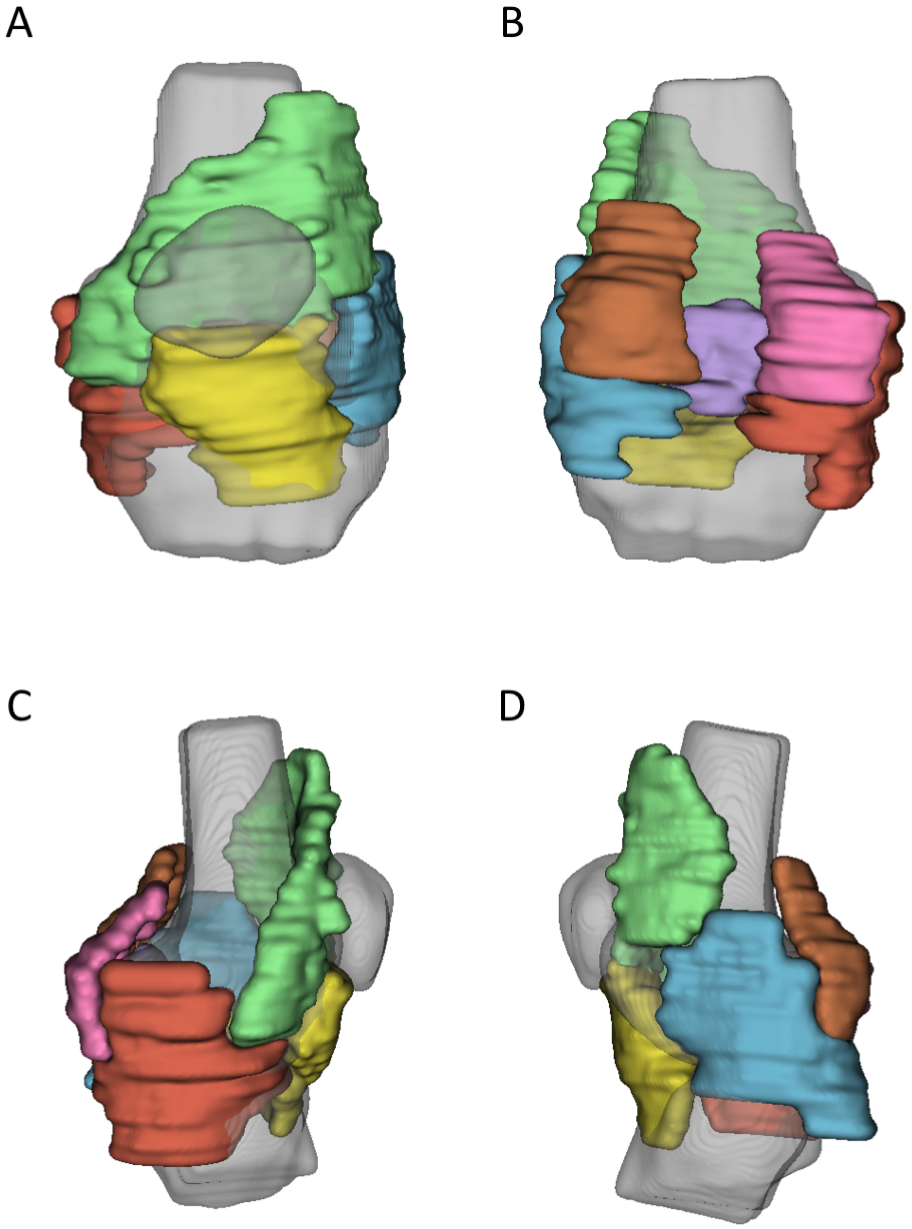
General details on the acquisition of MR data for the AMROA study are described in chapter 5 (section 5.2.3, Table 5.2). Sequences used for DCE-MR included pre-contrast variable flip angle T1 mapping sequences and a multi-phase dynamic sequence (temporal resolution for entire volume 14s). At the end of the sixth phase of the dynamic sequence, gdCA (Dotarem®, Guerbet, Paris, France) was administered at a dose of 0.2 mM/kg and a rate of 3 mL/s followed by a 50 mL saline chaser at the same rate. Imaging continued for a further 29 phases (35 phases total, acquisition time 8 min 10 s) following the gdCA administration. Pre and post gdCA 3D fat-saturated SPGR sequences were also obtained for synovial segmentation purposes. The total time taken for all sequences used for DCE-MR analysis was approximately 21 minutes.

6.2.3 Image analysis – region of interest definition

Initial rough manual segmentation of the synovium was performed on the post-contrast 3D FS SPGR sequence by a radiologist, with definition of seven synovial regions of interest (ROIs): suprapatellar, Hoffa's fat pad, medial and lateral perimeniscal, intercondylar notch, medial and lateral posterior medial femoral condyles (Figure 6.1). Anatomical definitions of synovial ROIs are provided in Table 6.1.

Figure 6.1

3D rendering of synovial regions of interest with anterior (A), posterior (B), medial (C) and lateral (D) views. 3D rendering of femur, tibia and patella (grey) provided for reference. **ROI Key**: green – suprapatellar, yellow – Hoffa fat pad, red – medial perimeniscal, blue – lateral perimeniscal, purple – intercondylar notch, pink – posterior medial femoral condyle, orange – posterior lateral femoral condyle.



Next, enhancing voxels were defined by subtracting the pre-contrast 3D FS SPGR sequence from the matching post-contrast sequence using a shuffle transform(263). For a given voxel in the post-contrast image, the shuffle transform minimises the absolute difference between the signal intensity of that voxel and the corresponding voxel plus a defined neighbourhood (for this study the adjacent 3 x 3 voxels) in the pre-contrast image. This improves the quality of the subtracted images and is also robust to residual motion artefact following image registration (Figure 6.2).

Table 6.1

Anatomical definition of synovial regions of interest.

Region	Definition
Suprapatellar	From mid pole of patella to superior extent of suprapatellar bursa, including medial and lateral peripatellar recesses
Hoffa fat pad	From junction of patellar tendon/tibial tuberosity to mid pole of patella superiorly. Posteriorly extends to anterior aspect of intermeniscal ligament and anterior horns of medial/lateral menisci
Medial perimeniscal	Wraps around medial meniscus. Extends superiorly deep to medial collateral ligament (MCL) to level of femoral MCL origin. Extends inferiorly deep to MCL to tibial MCL attachment.
Lateral perimeniscal	Wraps around lateral meniscus. Extends superiorly deep to lateral collateral ligament complex to fibular collateral ligament femoral origin. Extends inferiorly to level of tibiofibular joint. Includes popliteus recess, if present.
Intercondylar notch	From superior aspect of anterior cruciate ligament (ACL) femoral origin to floor of intercondylar notch at intercondylar eminences. Posterior margin tangential to most posterior margin of posterior cruciate ligament, anterior margin at anterior aspect of ACL.
Posterior medial femoral condyle	From most proximal aspect of medial head of gastrocnemius tendon to superior border of posterior horn of medial meniscus. Extends to meet intercondylar notch region laterally and medial perimeniscal region medially. Includes semimembranosus/medial head of gastrocnemius bursa, if present.
Posterior lateral femoral condyle	From most proximal aspect of lateral head of gastrocnemius tendon to superior border of posterior horn of lateral meniscus. Extends to meet intercondylar notch region medially and lateral perimeniscal region laterally.

Figure 6.2

Example of the use of shuffle transform to improve quality of subtracted image compared to simple subtraction of registered images. A – pre-contrast 3D FS SPGR, B – post-contrast 3D FS SPGR, C – simple subtraction (following intensity-based registration), D – shuffle subtraction. Improved subtraction quality is seen when the shuffle transform is used.

The shuffle subtracted images were then converted to binary enhancing masks using Otsu thresholding(264). The intersection between this binary mask and the rough manual segmentation for each ROI was termed the 'volume of enhancing pannus' (VEP) following the naming convention established in RhA(263). The VEP mask was used for the extraction of DCE-MR QIB values for each synovial ROI and for the whole joint (all ROIs combined). In addition, the VEP mask was used to create an estimate of volume of synovial tissue (VEP volume, measured in mL) by multiplying the number of voxels included in the VEP mask by the voxel size. Similar pipelines have previously been used for estimating synovial volume in knee OA(265).

6.2.4 Image analysis – pharmacokinetic modelling

All DCE-MR images were registered using intensity-based registration to the post-contrast 3D SPGR FS sequence for the purpose of motion correction. This used an initial rigid transformation to bring images into approximate alignment followed by a window-limited deformable registration, performed using the Advanced Normalization Tools(266). The accuracy of this registration procedure was confirmed visually for every case.

A pre-contrast T1 map was created using a variable flip angle approach. The T1 value at each voxel was estimated from the signal at each flip angle α:

$$S = M_0 \sin(\alpha) \frac{1-\varepsilon}{1-\cos(\alpha)\varepsilon} \qquad \text{where } \varepsilon = \exp\frac{-TR}{T1} \tag{6.1}$$

Where M_0 is proportional to the estimated proton density of the tissue and MR gain settings and is also estimated during the fitting process. There was no correction for B1 (transmit field) inhomogeneity given that a local transmit coil was used, meaning that any effect was likely to be minimal.

The signal intensity data *S* from each voxel at each timepoint in the dynamic series was converted into a T1 value via rearrangement of equation 6.1. This was in turn converted into a concentration of gdCA *C* by comparing the observed T1 value to the native T1 value T_{10} given the known relaxivity r_1 of the gdCA (Dotarem®: 3.4 $s^{-1}.mM^{-1}$ at 3 T) by rearrangement of equation 6.2

$$\frac{1}{T_1} = \frac{1}{T_{10}} + r_1 C \tag{6.2}$$

This creates a time-concentration curve for each voxel in the dynamic series, which is used as an input to equation 6.3, the extended Tofts-Kety model(267):

$$C_t(t) = v_p C_p(t) + K^{trans} \int_0^t C_p(t') \exp\frac{-K^{trans}(t')}{v_e} dt' \tag{6.3}$$

Where $C_t(t)$ is the time-concentration curve for the voxel, $C_p(t)$ is the arterial input function (AIF), K^{trans} is the volume transfer coefficient between plasma and the extracellular extravascular space (units min^{-1}), v_e is the proportion of extracellular extravascular space, and v_p is the proportion of plasma.

A population AIF was used due to technical difficulties in creating an individual AIF at the knee, including small artery size and the achievable temporal resolution(268). All AIFs were corrected for individual haematocrit values(238).

The initial area under the time-concentration curve at 60s following the administration of gdCA ($IAUC_{60}$, units mM.s) was also considered as a candidate QIB.

6.2.5 Statistical analysis

Test-retest repeatability was assessed using baseline and 1-month whole joint data. Root-mean-square average coefficients of variation (RMSCV) were calculated for each QIB. The smallest detectable difference (SDD) was calculated for each QIB, representing the magnitude of change that would give 95% confidence of a change being genuine rather than due to measurement noise(269). To derive the SDD, correlation between the difference between the baseline and 1-month measurements and the mean value was assessed visually and quantified using Kendall's τ. If significant correlation was detected, this indicates that variability depends on the QIB value in which case SDD is expressed as a percentage change. In this case, the within subject coefficient of variation was used:

$$SDD = \sqrt{\frac{\sum_{i=1}^{N}\left(\frac{\sigma_i}{\mu_i}\right)^2}{N}} \times 1.96 \times \sqrt{2} \times 100\% \qquad (6.4)$$

If no significant correlation was detected, this implies that variability is independent of QIB value in which case SDD is expressed as an absolute value. In this case, the within subject variance was used:

$$SDD = \sqrt{\frac{\sum_{i=1}^{N}{\sigma_i}^2}{N}} \times 1.96 \times \sqrt{2} \qquad (6.5)$$

Discrimination between OA and HV participants was assessed using baseline data. Descriptive statistics were calculated for each group, and the standardised mean difference (with 90% confidence intervals) was estimated for each QIB.

Six-month and 1-year changes in each QIB were assessed using descriptive statistics. ANCOVA was used to estimate the adjusted mean 6-month and 1-year change in each QIB in each group using baseline measurement as a covariate to account for regression to the mean(270). The number of participants with changes in each DCE-MR QIB greater than the SDD was calculated.

All statistical analyses were performed in RStudio version 1.0.143.

6.3 RESULTS

6.3.1 Participants

Participant characteristics are as described in chapter 5 (section 5.3.1). One original cohort OA participant declined gdCA administration at the six-month visit and did not attend the 1-year visit, therefore was not included in 6-month or 1-year change analyses.

6.3.2 Test-retest repeatability

RMSCV and SDD values for each parameter are provided in Table 6.2. The variability of K^{trans}, $IAUC_{60}$, VEP volume was not significantly correlated with the value of the parameter, so absolute SDD values were calculated. Variability of v_e was significantly correlated with parameter values, so percentage SDD values were calculated. There were a large number of 0 values for v_p, meaning calculation of RMSCV and SDD values were not possible. This parameter was not used for further analyses. Due to the very high RMSCV for v_e and the presence of implausible values (e.g. v_e greater than 1), this parameter was also not used for further analyses.

Table 6.2

Root-mean-square average coefficients of variation and smallest detectable changes for DCE-MR QIBs. The mean baseline value of each biomare in OA participants is also given to provide context to the SDD.

Parameter	RMSCV(%)	SDD*	OA mean (SDD as %)
K^{trans} (min^{-1})	17.8	0.013	0.041 (32)
$IAUC_{60}$ (mM.s)	32.3	6.74	8.98 (75)
VEP volume (mL)	34.8	71.5	69.2 (103)
v_p	NA	NA	0.0015 (NA)
v_e	61.2	169.5%	0.93 (NA)

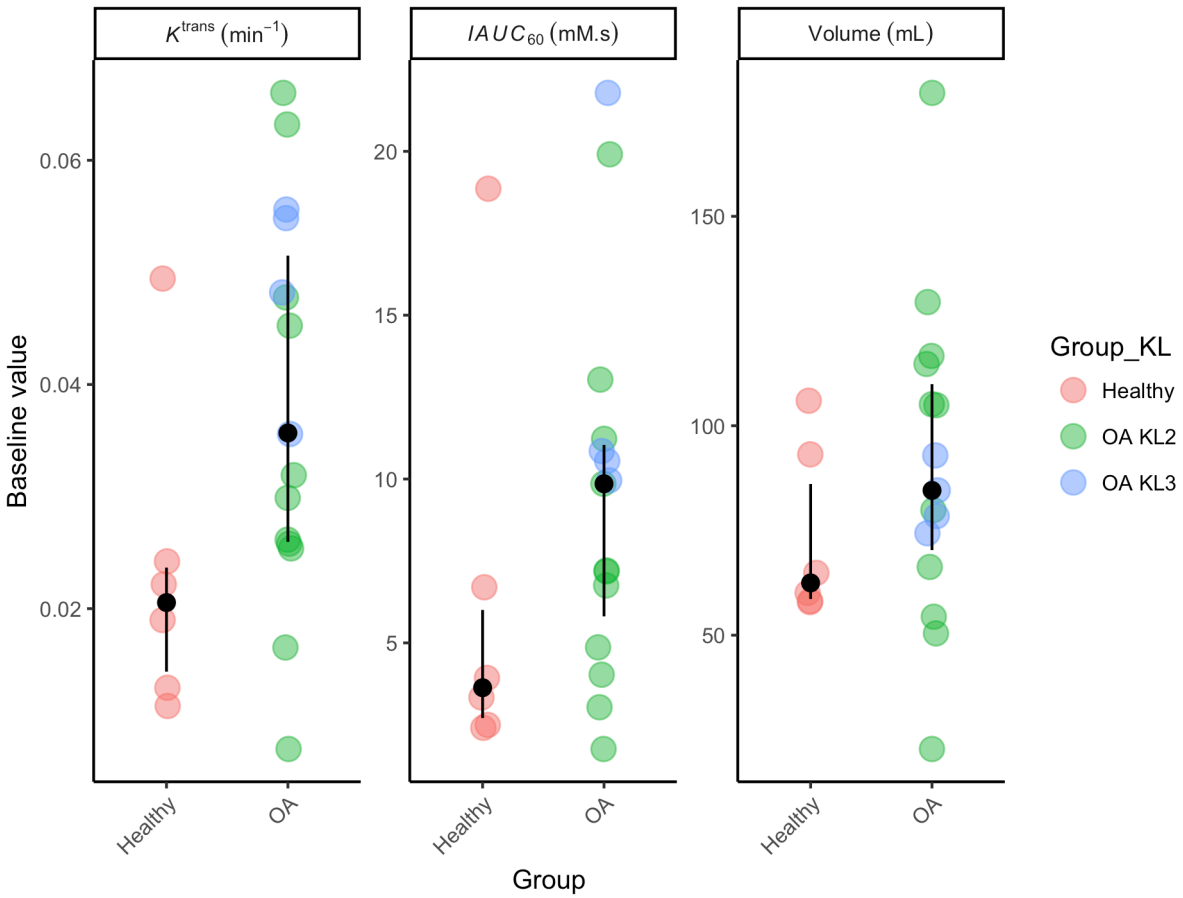
*provided as absolute values unless otherwise stated. **Abbreviations:** NA – not applicable

6.3.3 Baseline between-group differences

Baseline between group differences for the whole joint are illustrated in Figure 6.3. Plots for individual ROIs are provided in Appendix 6.1. One HV participant had much higher values of K^{trans} and $IAUC_{60}$ when compared to other HV participants across all ROIs. On further investigation, it was determined that this HV had taken part in karate practice the night before each study visit, and also had an undisclosed history of gout (never having affected the knee). Possible explanations considered for this value were that this represented part of the normal range of healthy values, or that the presence of gout or recent intense physical activity had confounded measurement.

Figure 6.3

Baseline DCE-MR QIB values. Black dots are median values, with interquartile range error bars.



The standardised mean difference (SMD) between OA and HV groups was 0.94 (90% confidence interval 0.11 to 1.77), 0.54 (-0.26 to 1.35) and 0.50 (-0.31 to 1.30) for K^{trans}, $IAUC_{60}$ and VEP volume respectively. Excluding the outlier HV case, SMDs were 1.34 (0.42 to 2.25) for K^{trans} and 1.12 (0.23 to 2.02) for $IAUC_{60}$. Analysis of synovial ROIs

revealed the highest between group differences for the intercondylar notch and medial and lateral perimeniscal ROIs for K^{trans} and $IAUC_{60}$ (Appendix 6.1). Between group differences were smaller for VEP volume for all synovial ROIs, with the largest difference for the suprapatellar ROI.

6.3.4 6-month and 1-year changes

Changes in DCE-MR QIBs over the duration of the AMROA study are summarised in Figure 6.4, with data for all synovial ROIs provided in Appendix 6.2. The mean 6-month and 1-year changes in DCE-MR QIBs, adjusted for baseline value, are provided in Table 6.3. All 90% confidence intervals for 6-month and 1-year mean changes in QIBs included 0 (no change) for both groups with the exception of 1-year change in VEP volume in HV.

Table 6.3

Adjusted mean 6-month and 1-year changes in DCE-MR QIBs.

Parameter	**6-month LSmean Δ (90% CI)**		**1-year LSmean Δ (90% CI)**	
	OA (n = 12)	**HV (n = 5)**	**OA (n = 9)**	**HV (n = 4)**
K^{trans} (min^{-1})	0 (-0.011, 0.010)	0 (-0.016, 0.014)	0 (-0.01, 0.01)	0.01 (-0.01, 0.03)
$IAUC_{60}$ (mM.s)	0.3 (-1.8, 2.4)	-1.5 (-4.5, 1.4)	1 (-2.0, 4.0)	1.7 (-2.8, 6.2)
VEP volume (mL)	26 (-2, 54)	28 (-12, 67)	16 (-2, 35)	34 (6, 62)

Abbreviations: LS mean – least-squares mean, i.e. mean change value adjusted for baseline value via ANCOVA

For K^{trans}, 5 out of 12 OA and 1 out of 5 HV participants had changes greater in magnitude (regardless of direction) than the SDD at 6-months. One out of 9 OA and 2 out of 4 HV participants had changes greater than the SDD at 1-year. Corresponding values for $IAUC_{60}$ and VEP volume are provided in Table 6.4.

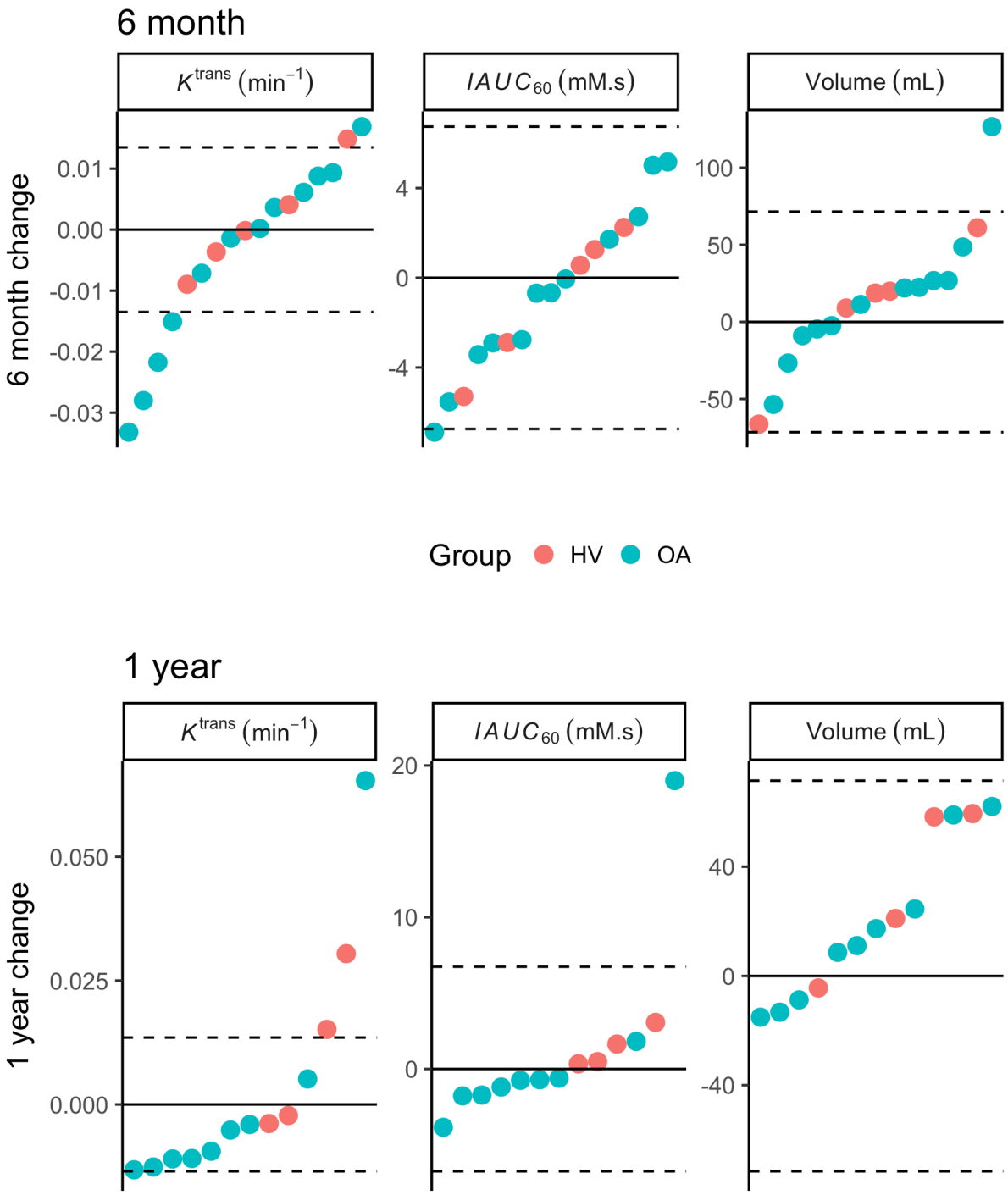
Table 6.4

Number of participants with 6-month and 1-year changes greater in magnitude than the SDD.

Parameter	**Number of participants with changes > SDD**			
	At 6-months		At 1-year	
	OA (n = 12)	HV (n = 5)	OA (n = 9)	HV (n = 4)
K^{trans} (min^{-1})	5	1	1	2
$IAUC_{60}$ (mM.s)	1	0	1	0
VEP volume (mL)	1	0	0	0

Figure 6.4

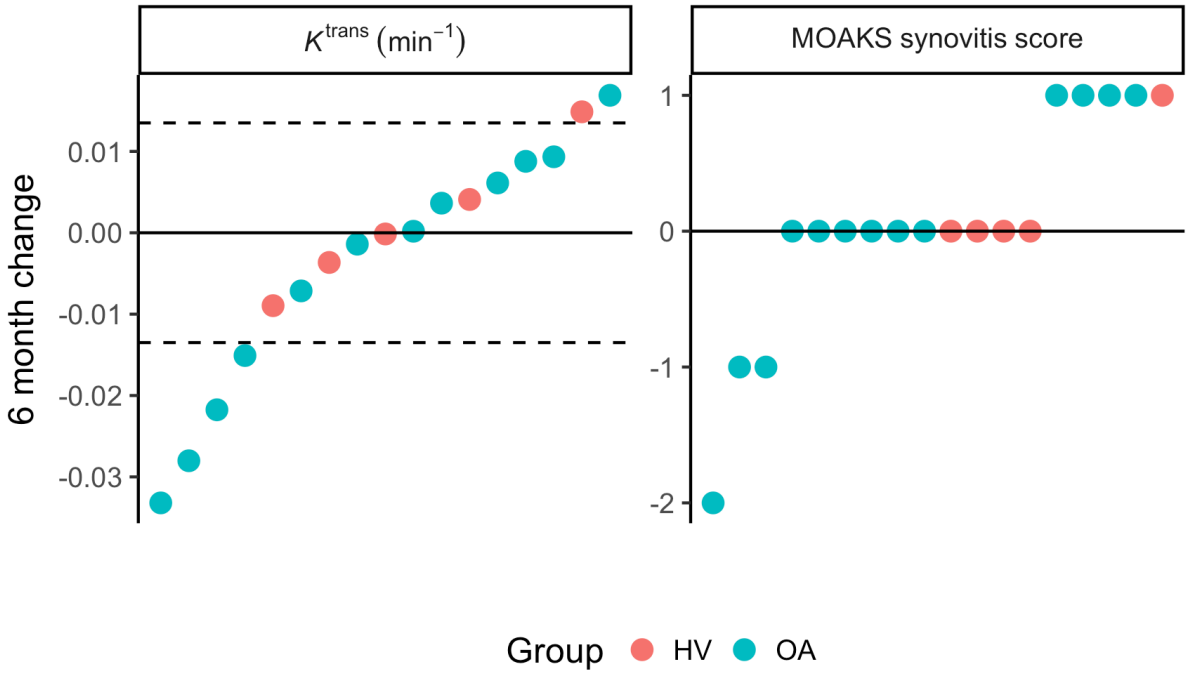
6-month and 1-year changes in DCE-MR QIBs. Each dot represents an individual participant. Dotted lines are ±SDD values.



A comparison of 6-month changes in K^{trans} and MOAKS synovitis score (sum of effusion synovitis and Hoffa synovitis scores, see chapter 5 for detail) is provided in Figure 6.5.

Figure 6.5

Individual participant 6-month changes in K^{trans} and MOAKS synovitis score. Dotted lines for K^{trans} are ±SDD values.



There was little concordance between participants with changes in K^{trans} and participants with changes in MOAKS synovitis score (Table 6.5). For example, one participant had an increase in K^{trans} but a reduction in MOAKS synovitis, and one participant had a decrease in K^{trans} but an increase in MOAKS synovitis.

Table 6.5

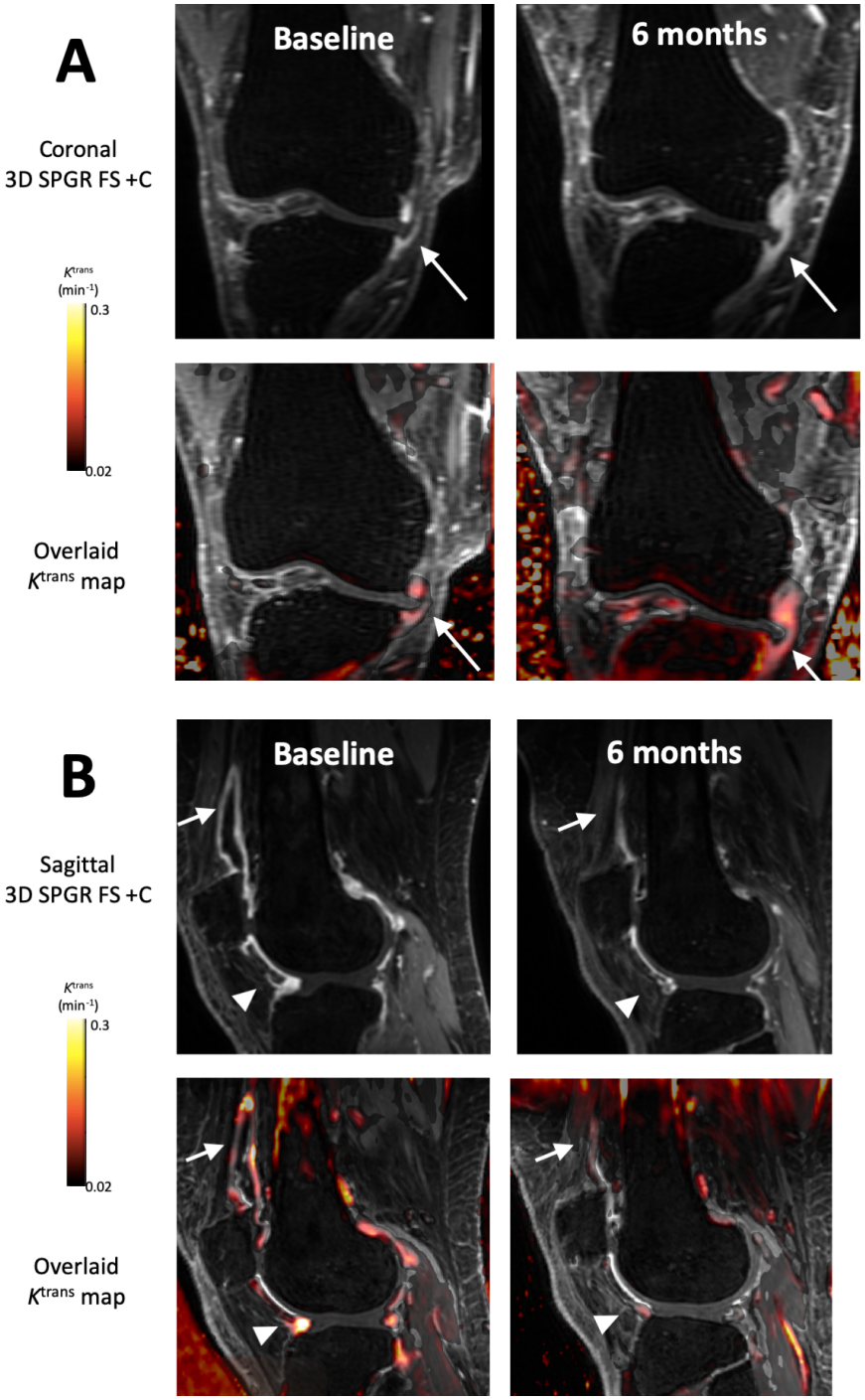
Confusion matrix demonstrating concordance between changes in K^{trans} and MOAKS synovitis score. Note total values are not provided (bottom right) due to different numbers of participants being assessed by each method (one OA participant declined contrast at 6-month visit).

		MOAKS synovitis			
		Increase	Decrease	No Δ	Total
K^{trans}	Increase	0	1	1	2
	Decrease	1	1	2	4
	No Δ	4	1	-	-
	Total	5	3	-	-

Example images of participants with significant changes at 6-months are provided in Figure 6.6.

Figure 6.6

Example images from participant with significant increase (A) and decrease (B) in K^{trans} at 6-months. In (A), note extruded medial meniscus with cuff of surrounding synovitis (white arrow). At 6-months, the synovitis has increased both in amount and intensity. In (B), not distention of suprapatellar pouch (arrow) and synovitis adjacent to anterior horn of lateral meniscus (arrowhead) at baseline, with marked reduction at 6-months.



6.4 DISCUSSION

This study suggests that K^{trans} is the optimum DCE-MR QIB for use in experimental medicine studies, with the best test-retest reproducibility, best discrimination between OA and HV participants, and greatest responsiveness as judged by the number of participants showing detectable changes from baseline over 6-months and 1-year.

Data on test-retest repeatability of DCE-MR QIBs in OA have not previously been published. Coefficients of variation are similar to previous DCE-MR studies in both musculoskeletal and non-musculoskeletal applications(261,271,272). Repeatability appears to be acceptable for experimental medicine studies given the finding of participants with changes exceeding the SDD at 6-months and 1-year. The test-retest repeatability interval chosen for the AMROA study (1-month) is relatively long when compared to the time over which fluctuations in synovitis are known to occur in OA. Therefore, 'true' methodological test-retest variability, excluding biological variation, is likely to be lower than the variability described here.

DCE-MR has also not previously been used to compare individuals with knee OA with healthy controls in the published literature. Higher values of K^{trans}, $IAUC_{60}$ and VEP volume were found in OA participants as would be expected. The highest point estimate for between-group SMD was for K^{trans}, albeit with overlapping confidence intervals with other QIB estimates. The results suggest that $IAUC_{60}$ is providing similar information to K^{trans}, which would be expected as both are driven by the initial gdCA uptake curve. However, K^{trans} seems to offer reduced variability, improved discrimination and improved responsiveness. One possible explanation for the observed superiority of K^{trans} is its improved physiological specificity when compared with $IAUC_{60}$. VEP volume differences between groups were also relatively large. It should be noted that no adjustment was performed for joint size. OA and HV groups were well matched for height and weight although not for sex (Chapter 5, section 5.3.1). More detailed adjustment for joint size via, for example, articular surface area is also possible but is difficult to interpret as surface area is also affected by the OA disease process.

One HV participant had much higher values of K^{trans} and $IAUC_{60}$ than did other HVs. As mentioned above, potential explanations for this outlier are the influence of

physical exercise the previous evening or subclinical synovitis due to gout or calcium pyrophosphate deposition arthropathy. The influence of prior exercise on DCE-MR parameters remains unknown and warrants further investigation given that this is a potential confound. If prior exercise is indeed a source of unwanted variation, then it should be noted that the observation of unloading protocol (~45 minutes rest) for cartilage compositional imaging in the AMROA study has not removed the effect in this study. This suggests a longer period of prior rest may be required for DCE-MR.

There was a wide range of 6-month changes in DCE-MR QIBs in both positive and negative directions in OA participants. MOAKS synovitis scores also demonstrated greater changes at six months than did other semiquantitative parameters. This may reflect the fluctuating nature of synovitis in OA, which is well recognised clinically with the concept of OA flares(273). The variation described in this observational study should be borne in mind if using DCE-MR QIBs in further interventional studies, where it seems reasonable to assume that some participants will demonstrate 'significant' reductions in K^{trans} and $IAUC_{60}$ as part of the natural variation in these parameters rather than due to treatment effects. At a group level, adjusted mean changes for each QIB were close to 0 at six months and 1-year. An effective treatment targeting synovitis may be expected to reduce the adjusted mean difference at a group level. Using the standard deviation of change observed in this study, a reduction in mean K^{trans} of 0.015 min^{-1} over 6-months could be detected with 80% power with a sample size of 16 participants per group, assuming an active treatment vs placebo design and a type 1 error rate of 0.05. This would represent a reduction of about 35% from the OA baseline mean, and is a clinically feasible reduction based on previous studies of response to intra-articular corticosteroid administration(71).

Results of DCE-MR QIBs were relatively consistent between different synovial ROIs. However, baseline SMDs for K^{trans} and $IAUC_{60}$ were highest in the perimeniscal and intercondylar notch regions. These are not areas which are scored by MOAKS or other commonly used whole joint scoring systems, highlighting a possible shortcoming. Alternative semiquantitative scoring systems for synovitis do take these locations into account, although are less widely used(274).

The use of DCE-MR QIBs obtained from analysis of dynamic data provides substantial additional information for assessment of synovitis by quantifying intensity of

inflammation. Morphological measures (such as VEP volume) can quantify the amount of synovial tissue, but not the intensity of inflammation. As well as playing a role in assessing response to treatment, 'intensive' (vs 'extensive') biomarkers of synovitis may have an important role to play in gaining better insight into the OA disease process. Evidence from animal and *ex-vivo* studies suggests that synovitis plays an important role in promoting adjacent cartilage degradation and stimulating marginal osteophyte formation(25,275). QIBs such as K^{trans} may offer a novel way of examining such hypotheses in *in-vivo* human studies of OA onset and progression.

The extended Tofts model used for pharmacokinetic modelling in this study was originally developed for DCE-MR of solid tumours. This study agrees with previous work highlighting the shortcomings of this model for synovitis(261). The model assumes that the contrast agent concentration within a voxel of interest will reach a plateau with possible wash-out phase. Voxel-wise analysis of uptake curves in this study's data demonstrate that in synovium, the plateau phase is not reached during the dynamic acquisition (approx. 8 minutes) which can lead to implausible estimates for v_e. The model also allows each voxel to contain a proportion of blood plasma volume (v_p), which is a sensible assumption in highly vascular tumours but perhaps less so in synovium where the vascular channels are much smaller. The DCE-MR parameters showing best performance in this study, K^{trans} and $IAUC_{60}$, are driven primarily by the initial uptake curve and therefore are unaffected by these assumptions.

The main study limitation is the use of a 1-month test-retest repeatability interval for DCE-MR QIBs which does not allow the isolation of methodological vs biological variability. The interval was chosen to assess test-retest repeatability for all QIBs included in the AMROA study, but this may have unfairly advantaged cartilage and bone QIBs which are less likely to show biological variability over this time. Recent concerns regarding deposition of gdCA in several tissues including the brain and bone marrow have reduced the ethical acceptability of repeat gdCA administrations over a short period of time, but a test-retest interval in the order of one week would likely be achievable(276). A second limitation is that the results presented are from a single centre and obtained with meticulous quality control, therefore extrapolation to multi-centre studies should be done with caution. However, previous work in other disease suggests that DCE-MR QIBs can be used in such a setting with appropriate training,

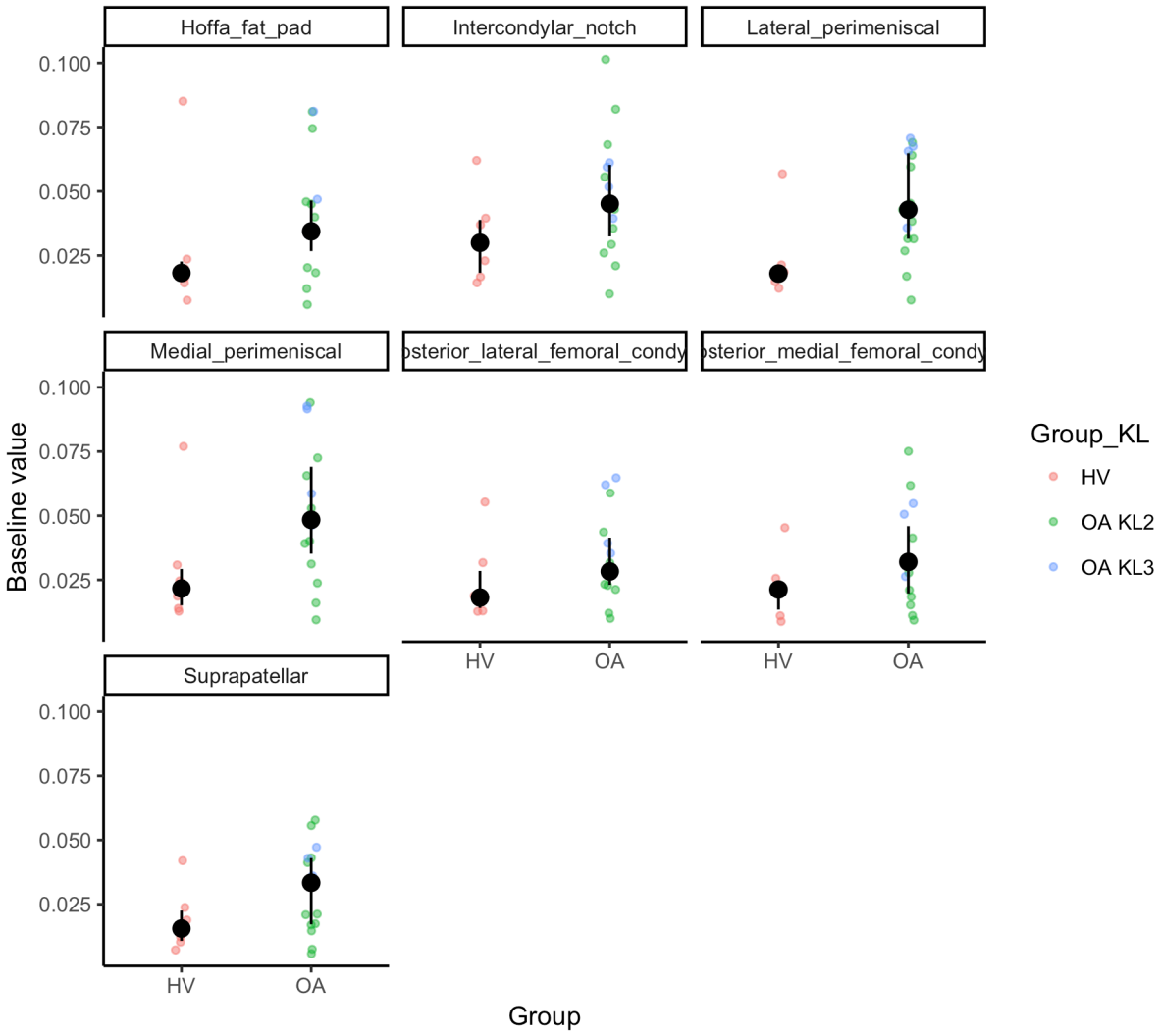
calibration and central image quality control(277). In particular, the use of a semi-automated pipeline as described in this study for defining the synovial ROI is likely to improve reliability in the multi-centre setting compared with alternative manual methods(278). A population-based arterial input function (AIF) was used for pharmacokinetic analysis of the DCE data. Such approaches improve reproducibility but may be less sensitive to change(268). The use of a population-based AIF was preferred as the temporal resolution of this study's volumetric acquisition (14 s/phase) precluded the accurate construction of individual AIFs.

In conclusion, this chapter has provided data on test-retest repeatability, discrimination between OA and 'normal' ageing and responsiveness for DCE-MR QIBs. K^{trans} appears to have the best performance across these domains and is therefore most likely to be useful in EM studies.

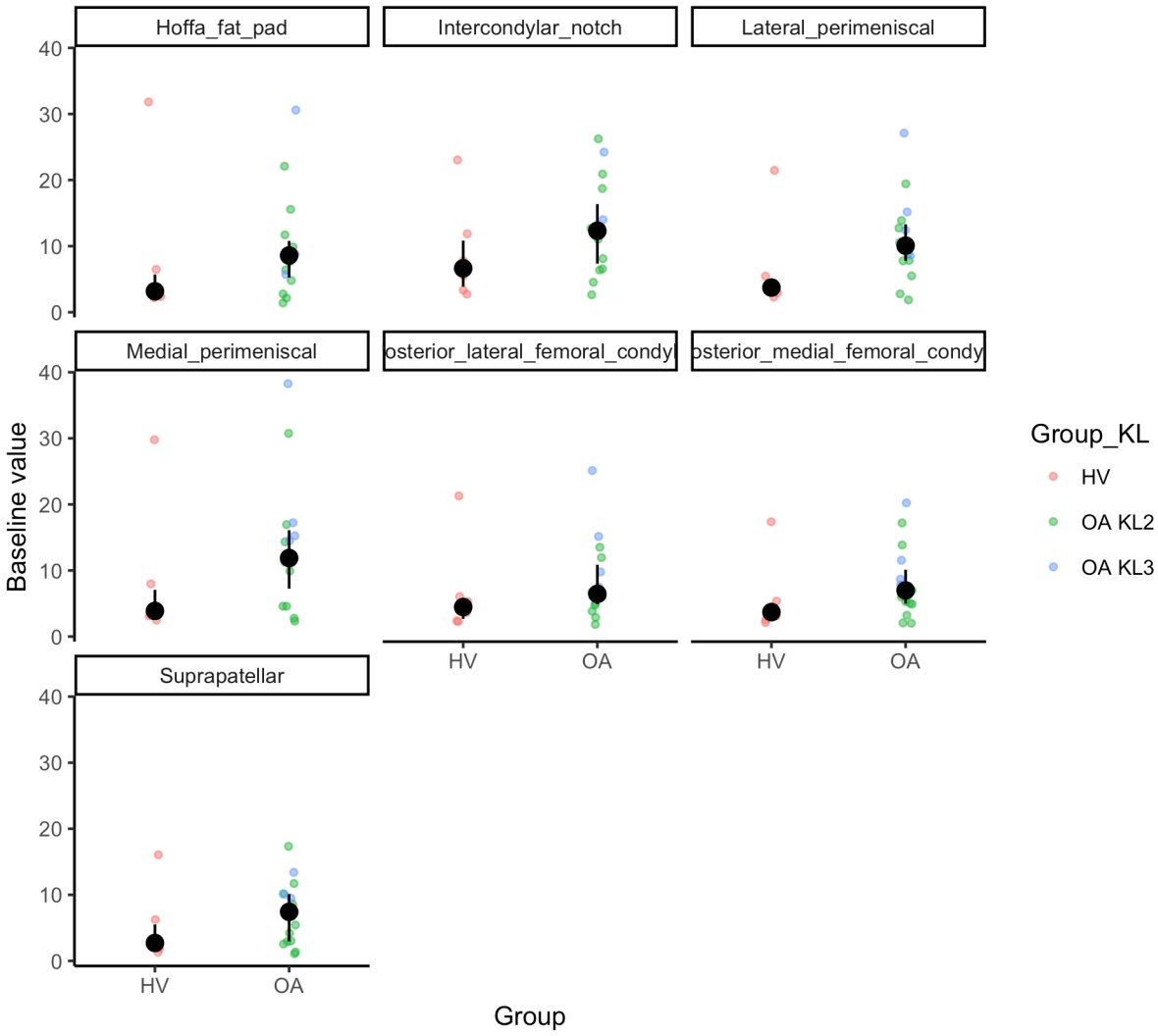
APPENDIX 6.1

Baseline DCE-MR QIB values in each synovial region of interest. Black dots are median values, with interquartile range error bars.

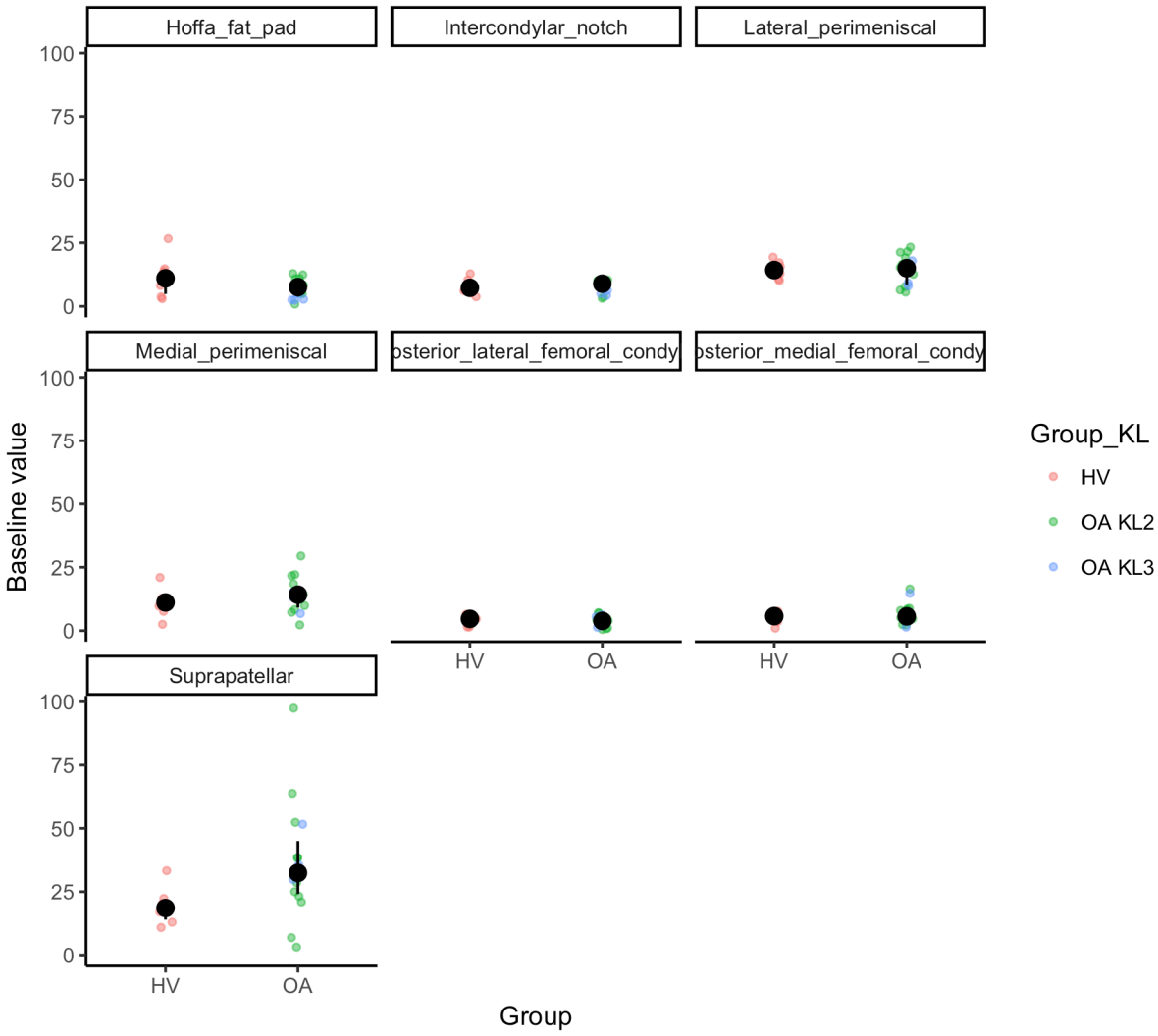
1. K^{trans} (min^{-1})



2. $IAUC_{60}$ (mM.s)


Hoffa_fat_pad
Intercondylar_notch
Lateral_perimeniscal
Medial_perimeniscal
osterior_lateral_femoral_condyl
osterior_medial_femoral_condy
Suprapatellar
Baseline value
Group
HV
OA
Group_KL
HV
OA KL2
OA KL3

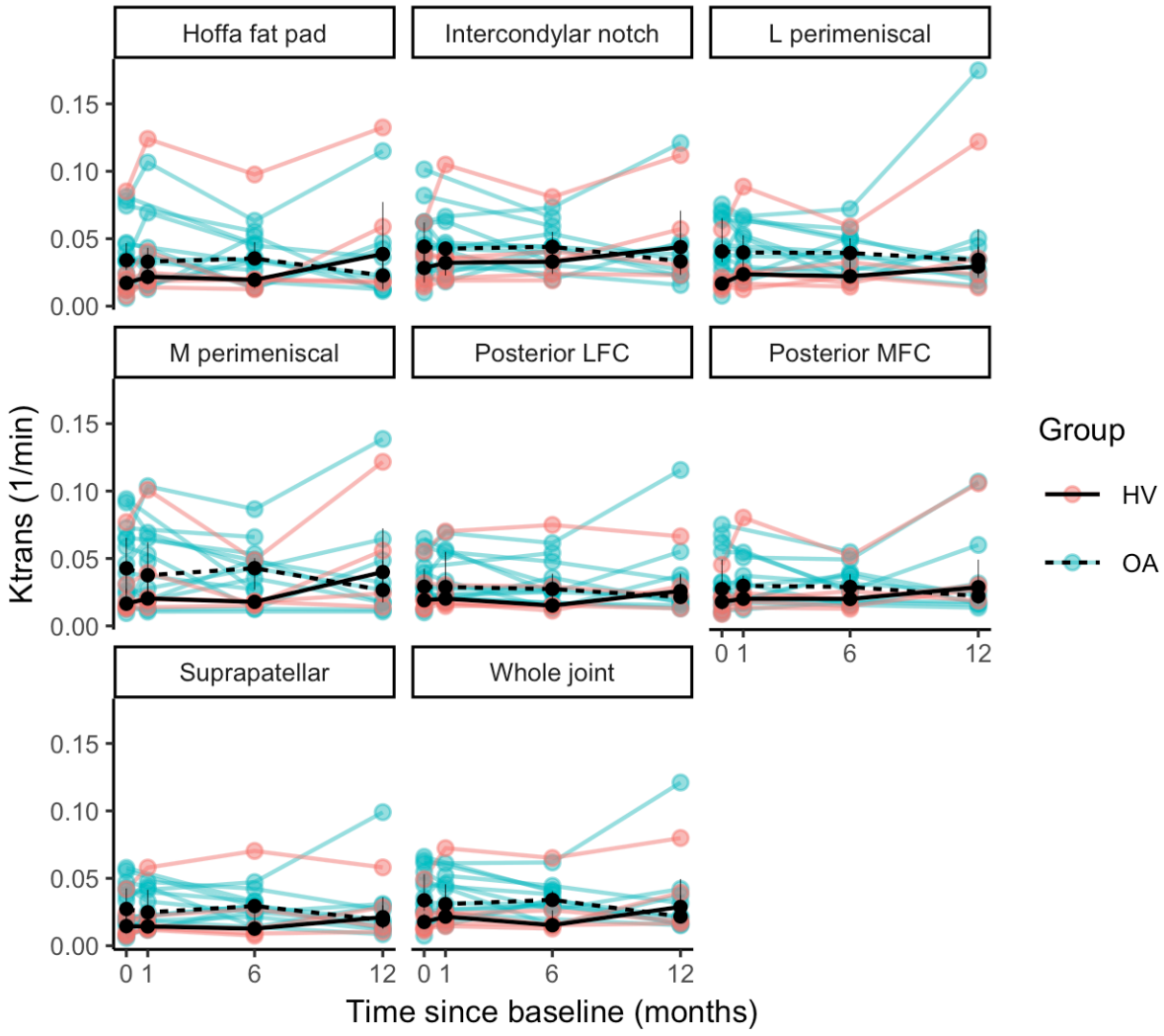
3. VEP volume (mL)



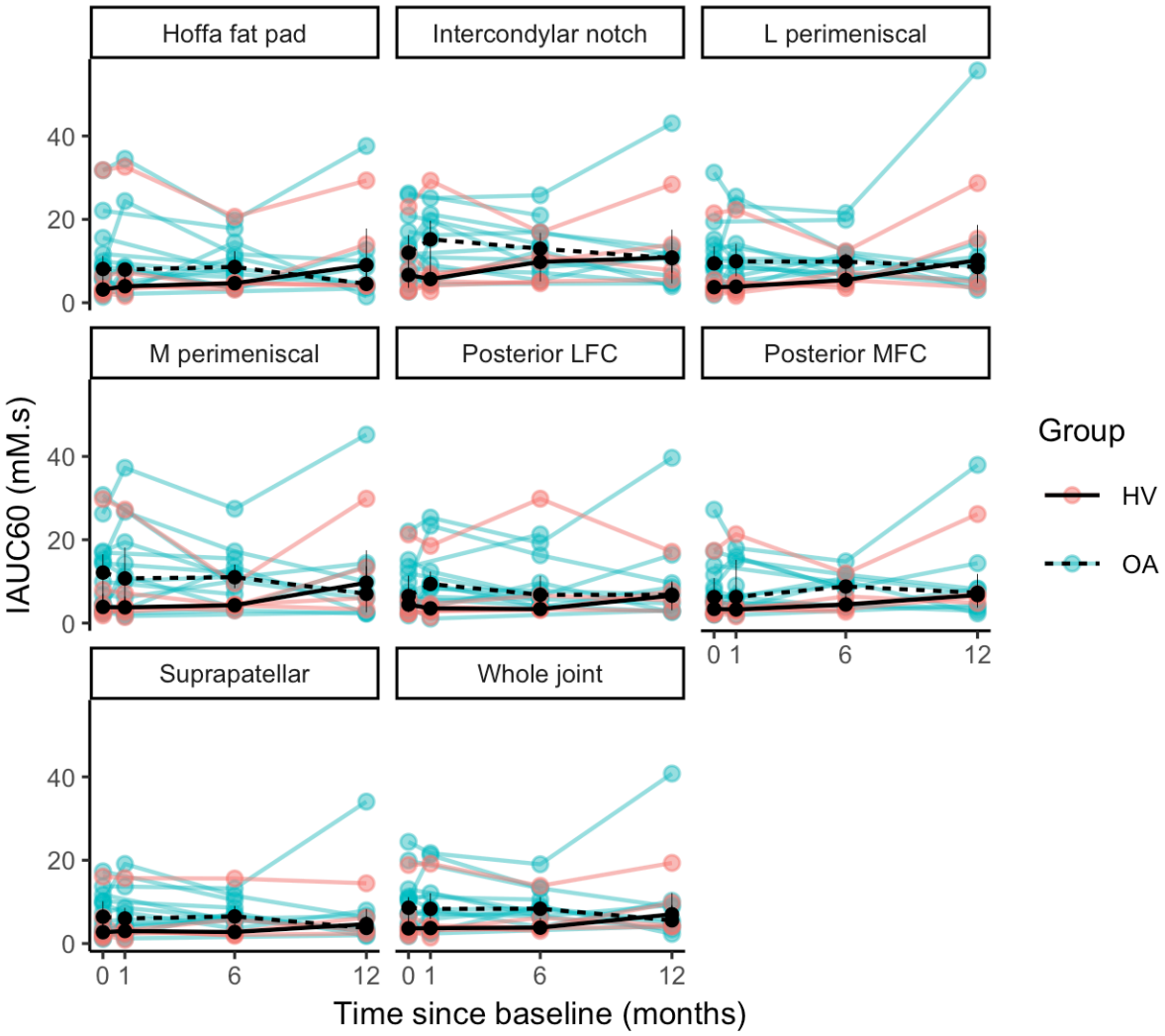
APPENDIX 6.2

Change in DCE-MR QIBs over study duration for each synovial ROI. Black dots are median values, with interquartile range error bars.

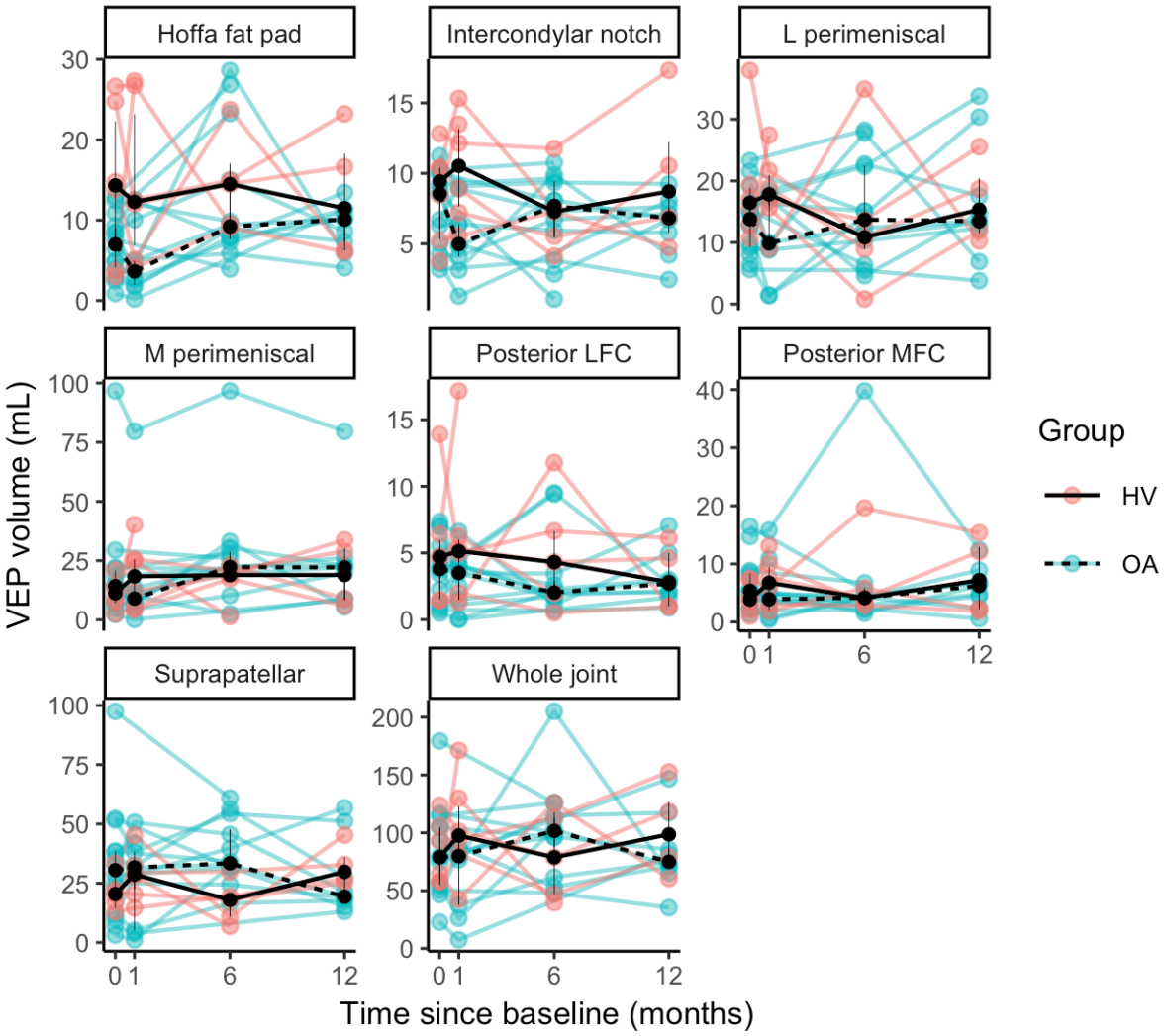
1. K^{trans}



2. $IAUC_{60}$


Hoffa fat pad
Intercondylar notch
L perimeniscal
M perimeniscal
Posterior LFC
Posterior MFC
Suprapatellar
Whole joint
IAUC60 (mM.s)
0
20
40
0 1 6 12
Time since baseline (months)
Group
HV
OA

3. VEP volume


Hoffa fat pad
Intercondylar notch
L perimeniscal
M perimeniscal
Posterior LFC
Posterior MFC
Suprapatellar
Whole joint
VEP volume (mL)
Time since baseline (months)
Group
HV
OA

CHAPTER 7:

MAGNETIC RESONANCE IMAGING OF SUBCHONDRAL BONE IN THE AMROA STUDY

Subchondral bone plays a critical role in OA onset and progression. Bone is a dynamic tissue, capable of remodelling in response to different loading conditions and endocrine and paracrine signalling. To evaluate the response of novel therapeutic agents on bone, sensitive imaging biomarkers of bone are desirable.

In this chapter I apply the texture analysis methodology described in chapter 4 to the AMROA study. I also analyse subchondral bone area which has been shown to be a sensitive marker of OA progression in previous studies. The results show that subchondral bone area measurement and subchondral bone texture analysis are highly repeatable with multiple participants demonstrating detectable changes at 6-months and 1-year. Subchondral bone texture is more responsive than subchondral bone area, but is more difficult to interpret from a biological or structural point of view.

7.1 INTRODUCTION

Subchondral bone plays an important role in OA development and progression(204). Because bone is a dynamic tissue, capable of remodelling in response to both mechanical and biological stimuli, there has been increasing interest in targeting potential disease modifying treatments (DMOATs) to this tissue(205,279). Sensitive imaging biomarkers of subchondral bone could help assess patient suitability for and response to putative DMOATs.

As discussed in chapter 4, multiple candidate quantitative imaging biomarkers (QIBs) for subchondral bone have been described using plain radiography, dual energy x-ray absorptiometry, computed tomography (CT), nuclear medicine and MR imaging(39,209,280,281). There are several reasons for preferring MR-based QIBs including the lack of ionising radiation and the ability to assess other tissues involved in the disease process in the same examination and determine the spatial relationship of changes in these tissues to changes in bone.

Two of the most promising MR QIBs for bone are subchondral bone area (SBA) and MR texture analysis (MR TA). SBA for both femur and tibia increases in OA(282,283). Change in SBA has been shown to discriminate patients with OA from controls and

has been associated with structural and symptomatic progression of OA in analyses of the Osteoarthritis Initiative (OAI)(63,226). Changes in SBA were detected over 3-months in a small OA cohort (n = 27) where there was no detectable change in cartilage thickness, demonstrating its potential utility as a QIB in experimental medicine studies(64). MR TA has been described previously in chapter 4 of this thesis. It has also shown the ability to discriminate patients with OA from healthy controls, and both baseline and change in MR TA are associated with structural OA progression(211,284). However, the test-retest repeatability and responsiveness to short-term change of MR TA remains uncertain. There has not been a previous head-to-head comparison of MR TA and SBA measurement to determine which has the most desirable QIB profile for use in experimental medicine studies.

Therefore, the purpose of this study was to evaluate the performance characteristics of two candidate MR QIBs of subchondral bone, subchondral bone area and subchondral bone texture, for use in experimental medicine (EM) studies.

7.2 METHODS

7.2.1 Participants & Ethical Approval

Ethical approval, participant recruitment and eligibility criteria are described in chapter 5 (sections 5.2.1 and 5.2.2).

7.2.2 Image Acquisition

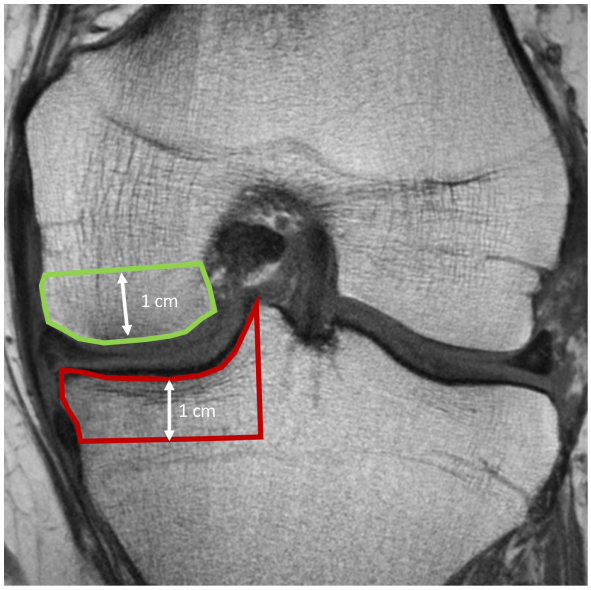
Details on the acquisition of MR data for the AMROA study are described in chapter 5 (section 5.2.3, Table 5.2). The coronal 2D T1-weighted fast spin echo (T1w FSE) sequence was used for subchondral bone texture analysis (Figure 7.1). The high spatial resolution pre-contrast sagittal 3D fat suppressed (FS) spoiled gradient echo (SPGR) sequence was used for segmentation of bone and cartilage from which SBA measurements were derived.

7.2.3 Image analysis – subchondral bone texture

The texture analysis methodology follows that described in chapter 4 (section 4.2.3). The three most central coronal-oblique images through the central medial tibiofemoral joint were identified with reference to localiser images and other sagittal sequences and used for subsequent analysis. Three central images were used here (compared to five central images used in chapter 4) as a 2D vs 3D sequence with thicker slices (2.5 mm vs 1 mm) and interslice gaps (2.5 mm vs 0 mm) was being used. The 2D T1w FSE sequence was chosen over the 3D gradient echo sequence used in the OAI bone ancillary study as previous work has demonstrated that it provides improved differentiation of osteoarthritic vs healthy bone, and also a fourfold shorter acquisition time(224).

Figure 7.1

Coronal 2D T1w FSE image through the central tibial plateau with example ROIs for tibial (red) and femoral (green) subchondral bone MR TA



The MR images were imported into a dedicated texture analysis program (MazDA v3.3, freely available at http://www.eletel.p.lodz.pl/programy/mazda/)(220). Regions of interest (ROIs) were created manually in the medial tibial and medial femoral subchondral bone on each coronal image. As in chapter 4, ROIs were defined superiorly and inferiorly by the bone-cartilage interface, medially and laterally by the margins of the tibial plateau and femoral condyle and extended for a depth of 1 cm into the subchondral bone (Figure 7.1).

The same 19 texture features as used in chapter 4 were calculated for each ROI with the same image compression settings and averaging of parameters across pixel offsets used. The values of each texture parameter on each of the three coronal images were averaged to give summary values for each parameter in each participant for medial tibial and medial femoral ROIs.

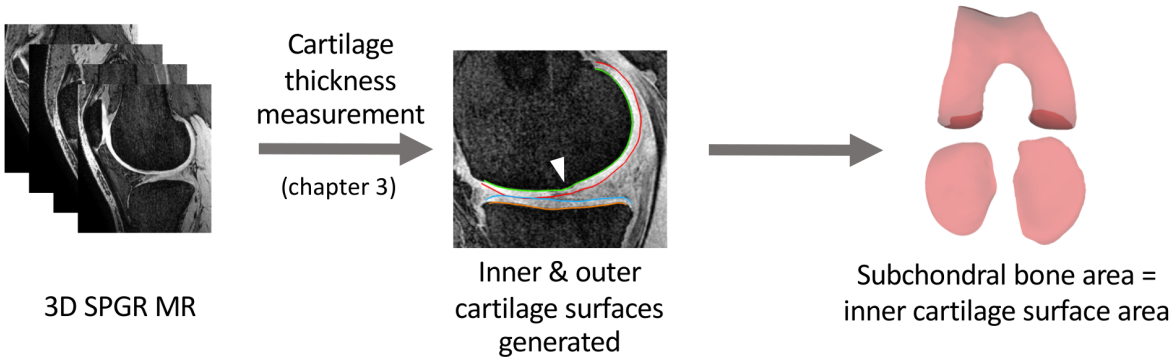
7.2.4 Image analysis – subchondral bone area

The 3D cartilage surface mapping (3D-CaSM) methodology described in chapters 3 & 8 yields measures of articular surface area for femur, medial tibia and lateral tibia. This is achieved via calculating the surface area of the inner cartilage surface mesh generated by the semi-automated thickness measurement process (Figure 7.2). The surface area calculation is performed in wxRegSurf (University of Cambridge Department of

Engineering, Cambridge, UK). In standard cartilage/bone morphology nomenclature, this would be defined as tAB (total area of subchondral bone) including osteophytes(231).

Figure 7.2

Subchondral bone area measurement pipeline. The cartilage thickness measurement pipeline described in chapters 3 & 8 generates accurate inner and outer cartilage surfaces. By definition, the inner surface corresponds to the subchondral bone contour. SBA is therefore simply the surface area of the generated inner surface meshes.



7.2.5 Statistical analysis

Test-retest repeatability was assessed using baseline and 1-month data. Root-mean-square average coefficients of variation (RMSCV) were calculated for SBA and MR TA features. The smallest detectable difference (SDD) was calculated for SBA for each surface as described in chapter 6 (section 6.2.5).

Due to the large number of MR TA features calculated (19), reduction in the dimensionality of the data was achieved via feature elimination. This involved excluding any texture features with suboptimal test-retest repeatability (RMSCV > 10%) from subsequent analyses. Individual texture features were then analysed separately in contrast to the composite texture score used in chapter 4. The reason for the different approach is the lower number of participants in AMROA which prevents reliable cross-validation and increases the likelihood of overfitting of any regression model. Simply using the same model as developed in chapter 4 to calculate texture score was also not appropriate as texture feature values will vary considerably between different MR sequences(285).

Discrimination between OA and healthy volunteer (HV) participants was assessed using baseline data. Descriptive statistics were calculated for each group, and the

standardised mean difference (with 90% confidence intervals) was estimated for SBA and MR TA features.

Six-month and 1-year changes in SBA and MR TA features were assessed using descriptive statistics. ANCOVA was used to estimate the adjusted mean 6-month and 1-year change in each QIB in each group using baseline measurement as a covariate. The number and proportion of participants with changes in each subchondral bone QIB greater than the SDC was calculated.

All statistical analyses were performed in RStudio version 1.0.143.

7.3 RESULTS

7.3.1 Participants

Participant characteristics are as described in chapter 5 (section 5.3.1). Two participants had 1-year visit images which were unsuitable for calculation of bone surface area due to motion artefact (1 participant) and radiographer error in scan prescription (1). Out of a total of 66 study visits across all participants, 11 visits were excluded from MR TA analysis. The two reasons for exclusion were poor image quality due to the tibial plateau being close to the inferior boundary of the knee coil sensitivity region - positioning of the knee in the coil for this study was lower than the positioning used in standard clinical practice so as to include the entirety of the suprapatellar region for DCE-MR analyses of synovitis (chapter 6) – and radiographer error in prescribing a standard clinical T1w sequence which has lower in-plane spatial resolution than the sequence used for MR TA. This meant that a total of 12 (7 OA/5 HV) participants contributed to test-retest assessment, 16 (11/5) contributed to 6-month change assessment, and 12 (8/4) contributed to 1-year change assessment of MR TA.

7.3.2 Test-retest repeatability

RMSCV and SDD values for SBA and MR TA features are provided in Table 7.1. Nine out of 19 MR TA features had RMSCV > 10% and were excluded from subsequent analyses.

7.3.3 Baseline between-group differences

Baseline between group differences in SBA and MR TA features are summarised in Figures 7.3 and 7.4. Mean SBA values were higher for all surfaces in the OA group with SMD (90% CI) of 0.52 (-0.29, 1.34), 0.53 (-0.29, 1.34) and 0.55 (-0.27, 1.36) for femur, medial tibia and lateral tibia respectively. The three tibial subchondral bone texture features demonstrating the best differentiation between OA and HV participants were the grey-level co-occurrence matrix (GLCM) parameters correlation (SMD [90% CI] = 1.26 [0.31, 2.20]) and entropy (0.53 [-0.36, 1.42]) and the grey-level histogram mean (-0.78 [-1.69, 0.12]). GLCM correlation (0.66 [-0.24, 1.55]) and entropy (0.50 [-0.39, 1.39]) were also two of the best three femoral subchondral texture features for

differentiating between OA and HV bone, the other being the run-length matrix (RLM) parameter run-length non-uniformity (0.38 [-0.50, 1.27]).

Table 7.1

Root-mean-square average coefficients of variation and smallest detectable changes for subchondral bone QIBs. The mean baseline value of each QIB in OA participants is also given to provide context to the SDD. For brevity, texture features which had RMSCVs > 10% (and were therefore not included in subsequent analyses) are not included.

Parameter	RMSCV(%)	SDD	OA mean (SDD as %)
Subchondral bone area			
Femur	2.1	249 mm^2	4986 mm^2 (5.0)
Medial Tibia	3.3	93 mm^2	1099 mm^2 (8.4)
Lateral Tibia	5.5	152 mm^2	1052 mm^2 (14.4)
Subchondral bone texture – tibia			
Mean	8.9	618	2528 (24.4)
Gr Mean	7.3	0.30	1.41 (21.3)
Gr NonZeros	2.2	0.05	0.86 (5.8)
Run fraction	1.1	0.03	0.88 (3.0)
SRLE	0.8	0.02	0.91 (2.2)
LRLE	3.3	0.13	1.45 (9.2)
RLNU	6.9	577	3404 (17.0)
Correlation	5.8	0.10	0.64 (15.0)
Entropy	2.6	0.18	2.51 (7.4)
Inv Df Mom	6.3	0.04	0.25 (16.8)
Subchondral bone texture – femur			
Mean	8.2	587	2748 (21.3)
Gr Mean	7.2	0.25	1.256 (20.2)
Gr NonZeros	3.0	0.07	0.84 (8.2)
Run fraction	1.5	0.04	0.87 (4.0)
SRLE	1.1	0.03	0.90 (3.0)
LRLE	4.3	0.18	1.48 (12.2)
RLNU	9.4	633	2793 (22.6)
Correlation	2.5	0.04	0.62 (5.8)
Entropy	2.8	0.18	2.40 (7.4)
Inv Df Mom	7.0	0.05	0.27 (19.8)

Abbreviations: Gr – gradient, SRLE – short run length emphasis, LRLE – long run length emphasis, RLNU – run-length non-uniformity, Inv Df Mom – inverse difference moment.

Figure 7.3

Baseline SBA values in each group. Black dots are median values, with interquartile range error bars.

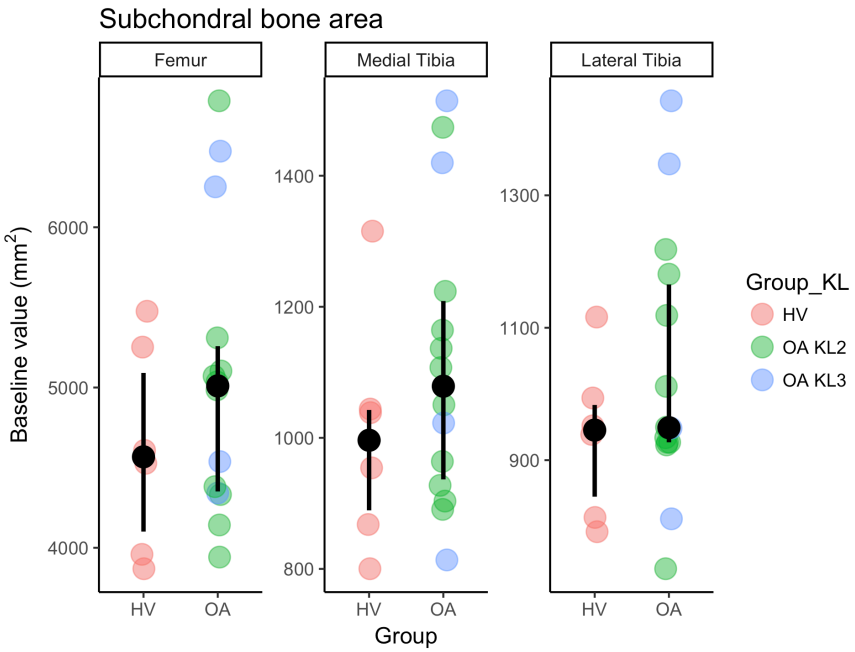
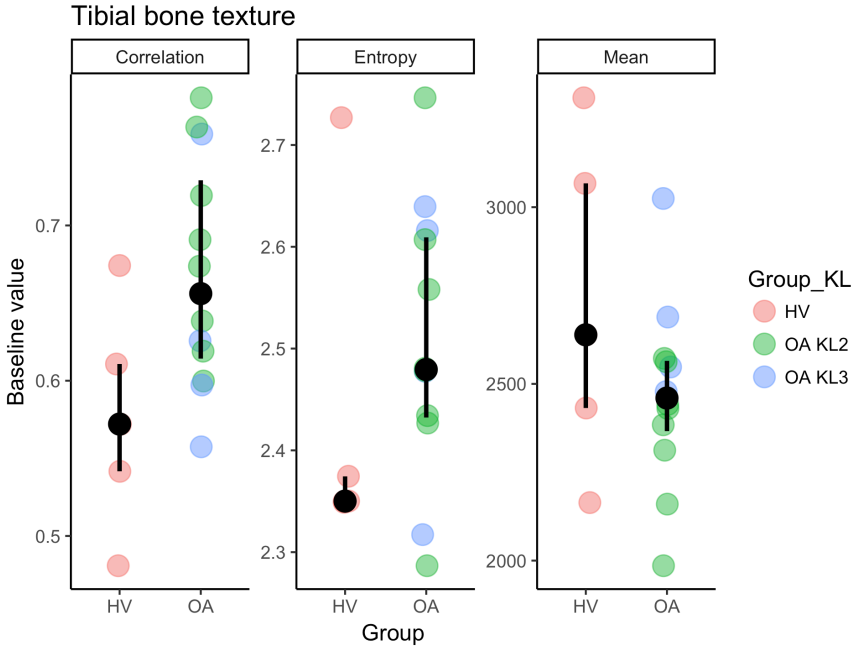
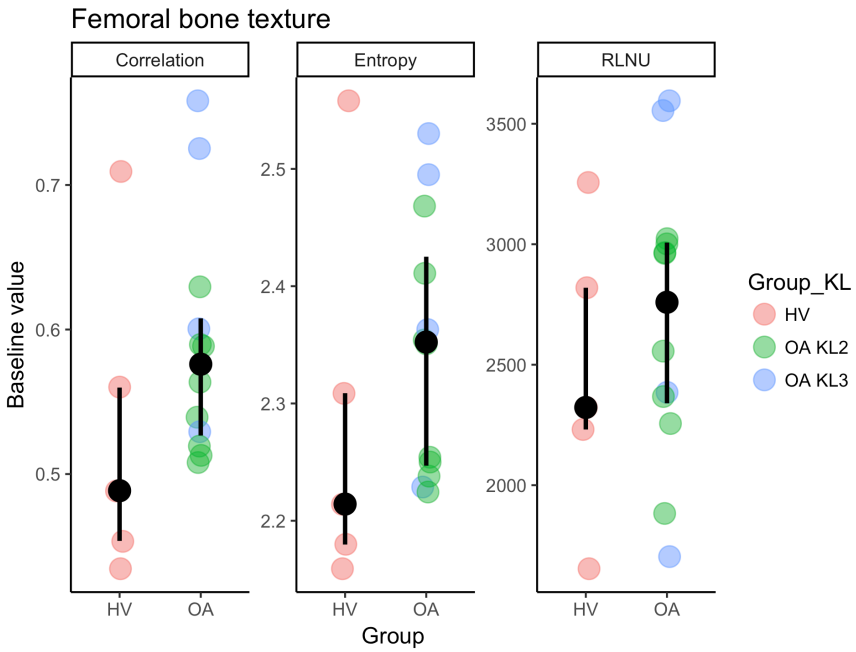


Figure 7.4

Baseline MR TA values in each group, limited to the three tibial and three femoral texture features which demonstrated the greatest standardised mean differences between OA and HV participants for brevity. Black dots are median values, with interquartile range error bars. **Abbreviations:** RLNU – run length non-uniformity





7.3.4 6-month and 1-year changes

The mean 6-month and 1-year changes in subchondral bone QIBs adjusted for baseline value are provided in Table 7.2. The lower 90% CI for mean femoral bone area increase over 1-year in the OA group was greater than 0. The lower 90% CI for mean medial tibial area increase in the HV group was also greater than 0.

Multiple MR TA features had mean 6-month and 1-year changes with 90% CIs excluding 0. The number of OA participants with 6-month and 1-year changes greater in magnitude than the SDD for subchondral bone QIBs is provided in Table 7.3. The parameters with most participants showing changes greater than the SDD at 6-months were the tibial texture features RLM short-run-length emphasis (SRLE) and long-run-length-emphasis (LRLE) and the femoral texture feature GCLM inverse difference moment (5 out of 11 OA participants and 2 out of 5 HV participants). At 1-year the femoral texture feature GLCM correlation had the highest number of participants demonstrating change greater than the SDD (7 out of 8 OA and 2 out of 4 HV participants) (Figure 7.5). Plots of individual changes for all SBA and MR TA parameters are provided in Appendix 7.1.

Figure 7.5

Dot plots and example images for most responsive subchondral bone QIBs at 6-months and 1-year. Dotted lines on dot plots represent +/- smallest detectable difference (SDD) limits. Images on left hand panel for participant with 6-month increase in tibial RLM long-run-length-emphasis (LRLE). Note focal area of homogeneous low signal (black arrowhead) in the tibial subchondral bone at both visits which represents a bone marrow lesion (BML). At the 6-month follow-up visit, the bone adjacent to this BML is of lower signal and is more homogeneous (black void arrowhead) which will lead to an increase in RLM LRLE. Images on right hand panel are for a participant with a 1-year decrease in femoral GLCM correlation. Note area of subchondral sclerosis present on both images (white arrowhead) but covering a larger area at the 1-year follow-up visit (white void arrowhead). More adjacent pixels having similar signal intensity will lead to an increase in GLCM correlation.

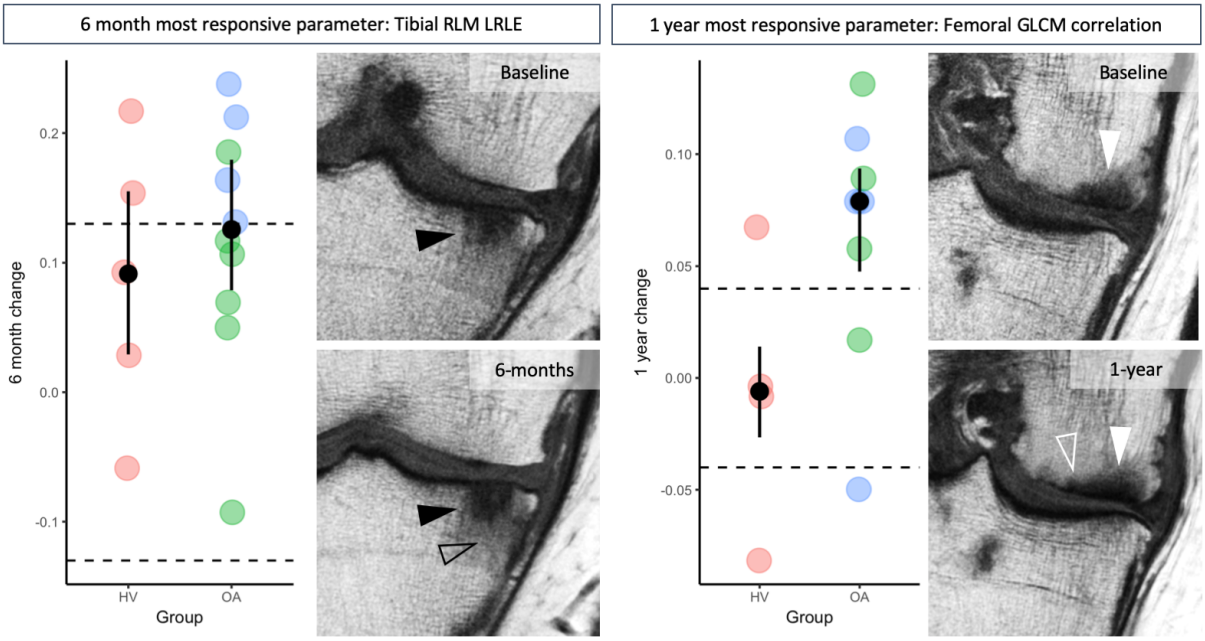


Table 7.2

Adjusted mean 6-month and 1-year changes in subchondral bone QIBs.

Parameter	6-month LSmean Δ (90% CI)		1-year LSmean Δ (90% CI)	
	OA	HV	OA	HV
Subchondral bone area (mm^2)				
	n = 13	n = 5	n = 7	n = 4
Femur	46 (-36, 128)	35 (-978, 168)	172 (116, 228)	-34 (-109, 40)
Medial Tibia	4 (-20, 29)	19 (-22, 59)	23 (-8, 54)	51 (9, 92)
Lateral Tibia	2 (-33, 37)	38 (-20, 95)	5 (-22, 31)	29 (-6, 65)
Subchondral bone texture - Tibia				
	n = 11	n = 5	n = 8	n = 4
Mean	-117 (-382, 147)	-61 (-442, 320)	-53 (-355, 249)	-88 (-527, 350)
Gr Mean	-0.2 (-0.3, -0.1)	-0.1 (-0.2, 0)	-0.3 (-0.4, -0.2)	-0.2 (-0.3, 0)
Gr NonZeros	-0.04 (-0.06, -0.03)	-0.03 (-0.05, 0)	-0.06 (-0.09, -0.04)	-0.03(-0.06, 0)
Run Fraction	-0.02 (-0.03,-0.01)	-0.02 (-0.03, 0)	-0.03 (-0.05, -0.02)	-0.02 (-0.04, 0)
SRLE	-0.02 (-0.03, -0.01)	-0.01 (-0.02, 0)	-0.03 (-0.04, -0.02)	-0.01 (-0.03, 0)
LRLE	0.12 (0.06, 0.18)	0.09 (0, 0.16)	0.18 (0.11, 0.25)	0.10 (0, 0.20)
RLNU	-421 (-675, -166)	-3 (-363, 357)	-661 (-811, -510)	-1222 (-1434, -510)
Correlation	0.05 (0, 0.10)	-0.02 (-0.09, 0.06)	0.07 (0.02, 0.12)	-0.02 (-0.09, .06)
Entropy	-0.08 (-0.14, -0.02)	-0.11 (-0.19, -0.02)	-0.13 (-0.22, -0.03)	-0.12 (-0.25, 0.02)
Inv Df Mom	0.03 (0.02, 0.04)	0.02 (0.01, 0.04)	0.05 (0.03, 0.07)	0.02 (0.01, 0.06)
Subchondral bone texture – Femur				
	n = 11	n = 5	n = 8	n = 4
Mean	-188 (-348, -29)	-75 (-303, 152)	-205 (-493, 83)	-176 (-587, 235)_
Gr Mean	-0.18 (-0.23, -0.12)	-0.14 (-0.22, -0.06)	-0.29 (-0.34, -0.24)	-0.25 (-0.33, 0.18)
Gr NonZeros	-0.05 (-0.07, -0.03)	-0.03 (-0.06, -0.01)	-0.08 (-0.10, -0.06)	-0.06 (-0.09, -0.03)
Run Fraction	-0.03 (-0.04, -0.02)	-0.02 (-0.03, -0.01)	-0.04 (-0.06, -0.03)	-0.03 (-0.05, -0.02)
SRLE	-0.02 (-0.03, -0.01)	-0.02 (-0.03, 0)	-0.04 (-0.04, -0.03)	-0.02 (-0.04, -0.01)
LRLE	0.14 (0.08, 0.20)	0.10 (0.02, 0.18)	0.24 (0.18, 0.30)	0.15 (0.07, .24)
RLNU	-434 (-575, -294)	-79 (-279, 121)	-555 (-724, -386)	-853 (-1093, -613)
Correlation	0 (-0.02, 0.02)	0.05 (0.01, 0.08)	0.06 (0.02, 0.11)	-0.01 (-0.07, 0.06)
Entropy	-0.13 (-0.18, -0.08)	-0.08 (-0.15, -0.01)	-0.17 (-0.22, -0.11)	-0.20 (-0.28, -0.11)
Inv Df Mom	0.04 (0.03, 0.05)	0.03 (0.01, 0.05)	0.06 (0.05, 0.08)	0.05 (0.03, 0.07)

Abbreviations: LS mean – least-squares mean, i.e. mean change value adjusted for baseline value via ANCOVA, otherwise as for Table 1

Table 7.3

Number of participants with 6-month and 1-year changes greater in magnitude than the SDD.

Parameter	Number of participants with changes > SDD at 6 months		Number of participants with changes > SDD at 1 year	
Subchondral bone area				
	n = 13	n = 5	n = 7	n = 4
Femur	3	0	2	0
Medial Tibia	5	0	1	1
Lateral Tibia	0	0	0	0
Subchondral bone texture - Tibia				
	n = 11	n = 5	n = 8	n = 4
Mean	1	2	1	2
Gr Mean	2	1	5	1
Gr NonZeros	4	1	5	1
Run Fraction	1	1	3	1
SRLE	5	2	4	2
LRLE	5	2	4	2
RLNU	4	2	5	3
Correlation	3	0	2	1
Entropy	2	2	3	1
Inv Df Mom	4	2	4	1
Subchondral bone texture – Femur				
	n = 11	n = 5	n = 8	n = 4
Mean	0	1	1	1
Gr Mean	3	1	5	1
Gr NonZeros	2	1	5	2
Run Fraction	1	1	4	1
SRLE	2	1	4	1
LRLE	3	1	3	1
RLNU	3	0	5	3
Correlation	4	2	7	2
Entropy	4	1	4	2
Inv Df Mom	5	2	6	2

Abbreviations: As for Table 7.1

7.4 DISCUSSION

Results from the AMROA study demonstrate that both SBA and MR TA have the potential to be useful QIBs of subchondral bone in experimental medicine studies. There are advantages and disadvantages to both. SBA is repeatable and demonstrates significant group level changes at 1-year, but is less responsive at an individual level. MR TA is both repeatable and responsive to change, but requires analysis of a number of different parameters whose relationship to the image being analysed can be difficult to interpret.

Repeatability of SBA assessment was good, particularly for femur and medial tibia which is where the majority of changes would be expected to occur in a medial OA population. The repeatability of SBA assessment in this study is comparable to most other previously published methods although worse than for a commercially available method which uses active appearance models (AAMs) and has test-retest coefficients of variation of less than 1% (64,282,283). One possible explanation for the better performance of the AAM-based method is that area is calculated for the central medial femoral area alone which has perhaps less potential for error than measuring the whole femoral area as performed in this study. Baseline between-group differences in SBA were in the direction that would be expected given previously described shape changes in OA(286). It should be noted that due to the small sample size, and to maintain consistency with other AMROA analyses, confounding variables such as sex and height were not adjusted for in the between group comparisons. At a group level, the average femoral SBA increased significantly in OA compared to HV participants over 1-year which is in keeping with previous work describing femoral area as more responsive than tibial area (11,20). At the individual level, relatively few participants demonstrated 6-month and 1-year changes greater than the SDC. Given the good test-retest repeatability demonstrated in the current work and previous studies, the main limiting factor in terms of responsiveness may be that the changes in bone area take place over a relatively long time period compared to other candidate QIBs.

MR TA features had variable repeatability, with the majority of features having test-retest RMSCVs of less than 10%. Interestingly, previous studies of MR TA (including chapter 4 of this thesis) have shown that the inter-observer reproducibility of MR TA

features is also variable(211,284). Clearly some texture features are very sensitive to small changes in the image or in ROI delineation making them less suitable for longitudinal analyses. Discrimination between groups was generally good, with the direction of difference in texture features between OA and HV groups in agreement with previous studies and generally in keeping with OA participants having more heterogeneous, less spatially organised subchondral bone. This was the first study to use MR TA in a prospective longitudinal setting. At an individual level, texture features had high responsiveness with multiple participants demonstrating changes greater than the SDC. At a group level, there were few differences between groups in average 6-month and 1-year changes.

One of the main disadvantages of MR TA is the limited interpretability of texture features. It is possible to describe approximately what texture features mean in terms of image heterogeneity and spatial organisation, but a more nuanced understanding of what parts of the image or ROI are driving the texture feature calculation is not possible. In this respect MR TA is something of a 'black box' methodology. While this does not prevent it from detecting and quantifying abnormality in subchondral bone, in the setting of EM studies it would be desirable to understand not just if a potential DMOAT is working but also some understanding of the underlying mechanism at a tissue level. MR TA does not provide an easy way to gain this understanding. Conversely, SBA is much easier to interpret and links directly to the underlying OA disease process. There is also a much larger body of work linking it to meaningful future clinical and structural outcomes(226,287). However, this has to be balanced against the reduced responsiveness compared to MR TA demonstrated in this study.

Both SBA and MR TA have the advantage over some other QIBs used in the AMROA study of requiring standard MR acquisitions (3D SPGR and 2D T1w FSE) which should be more easily translatable to the multi-centre setting when compared to QIBs which require more advanced acquisition methods such as DCE-MR (chapter 6) and cartilage compositional assessment (chapter 8). In addition, some of the most promising putative DMOAT data is for treatments targeting subchondral bone. For example, while large clinical trials of bisphosphonates have failed to demonstrate significant effects on patient reported outcome measures there has been some evidence of a slowing in structural progression particularly in certain subgroups(288,289). In

smaller trials, other therapeutics targeting subchondral bone such as oral salmon calcitonin and cathepsin K inhibitors have also shown promise(66,290). QIBs such as those used in this study may provide a way to both identify suitable participants for assessment of potential DMOATs targeting subchondral bone and provide surrogate endpoints for these studies.

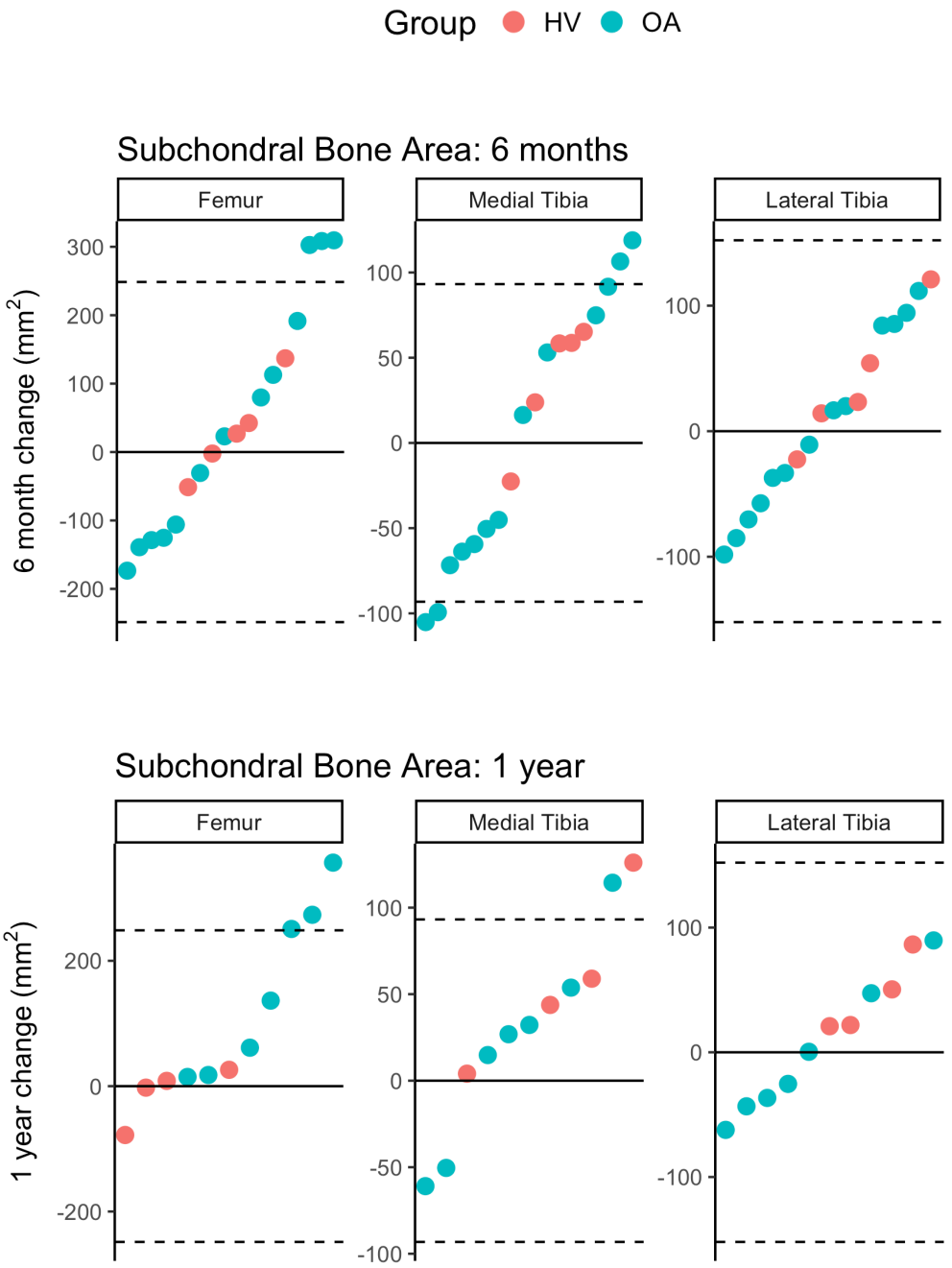
As well as being useful in EM studies, QIBs of subchondral bone could provide novel insights into OA pathophysiology. For example, measurement of SBA has led to the report of the so-called 'pie-crust' effect where there is a gradual spreading out of subchondral bone at the margins of the articular surface as OA develops. This is essentially describing osteophytosis in 3D and has debunked previous theories that osteophytes were only seen in weight bearing areas of the joint(63,204).

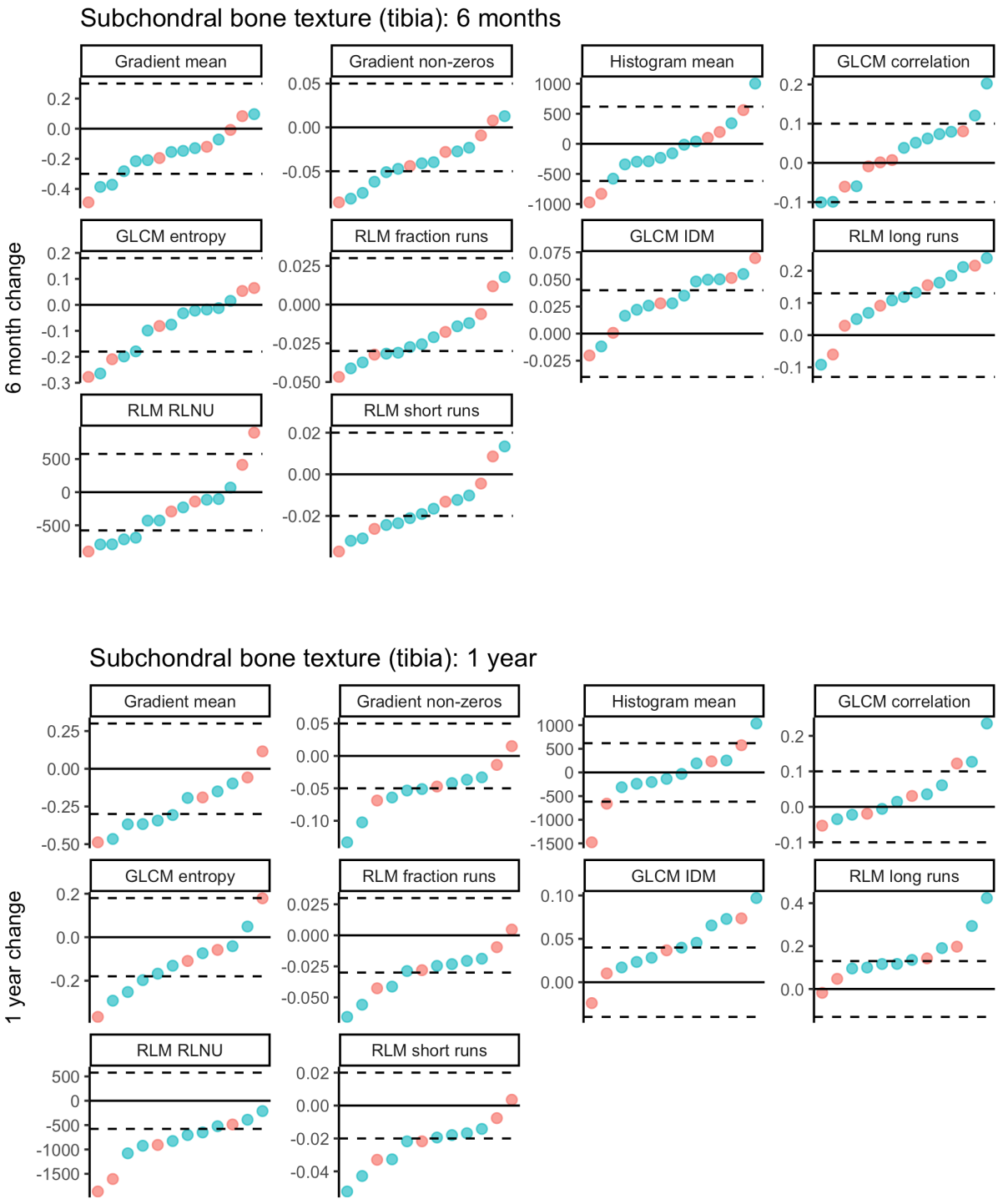
There are several study limitations. First, a relatively large number of coronal T1w images were not suitable for MR TA due to need to position knee lower than usual for DCE-MR acquisition and analysis. A study with a dedicated texture analysis acquisition protocol may yield improved repeatability, discrimination and responsiveness data meaning that the values provided here could be regarded as conservative. Second, so-called 'scanner drift' – the tendency for quantitative values to change with MR system software updates etc – cannot be excluded as a cause for change in texture features during the study. Quantitative metrics such as T2 and T1rho relaxometry can be adjusted for this using phantom based measurements. However, no standardised phantom for texture measurements exists and so this was not possible here. Finally, the assessment of SBA in this study required some manual interaction whereas some alternative methods (e.g. those using AAMs) are fully automatic. Although the test-retest repeatability and inter-observer reproducibility reported in this study are very good, it is likely that automation may improve this even further and therefore offer slightly improved responsiveness.

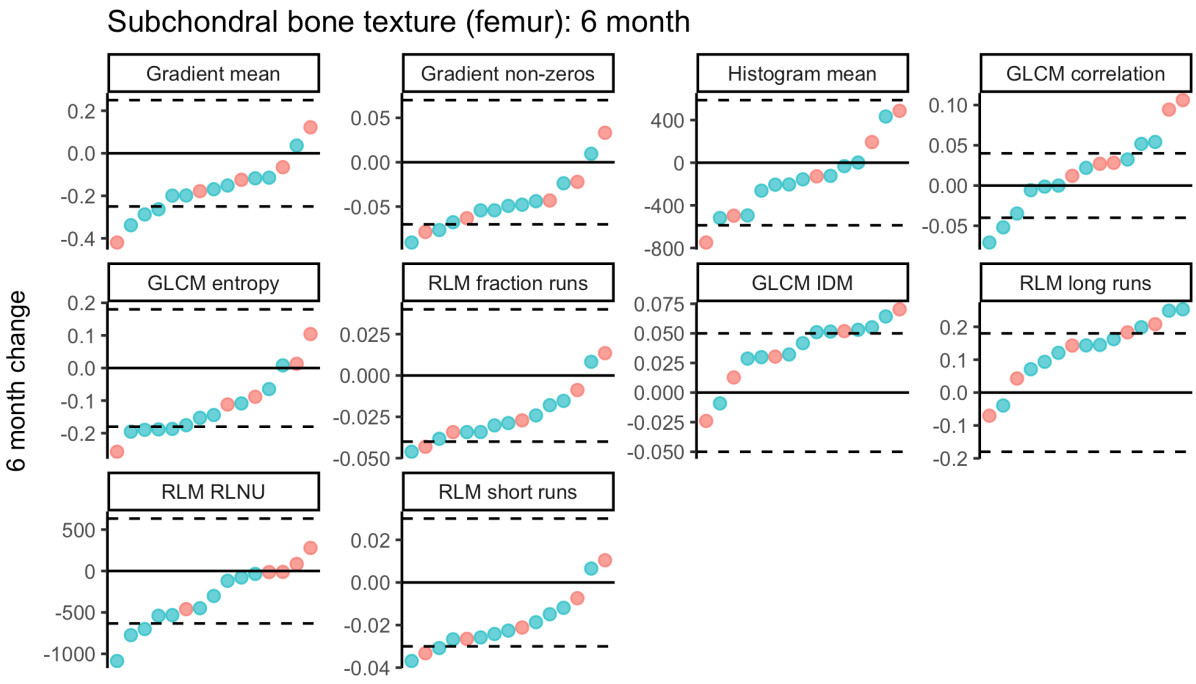
In conclusion, in this chapter I have assessed the likely utility of MR QIBs of subchondral bone in EM studies. Both MR TA and SBA are repeatable and able to distinguish between OA and HV subchondral bone. MR TA may have superior responsiveness to SBA, but is less interpretable in terms of how it relates to the OA disease process which may hinder its uptake.

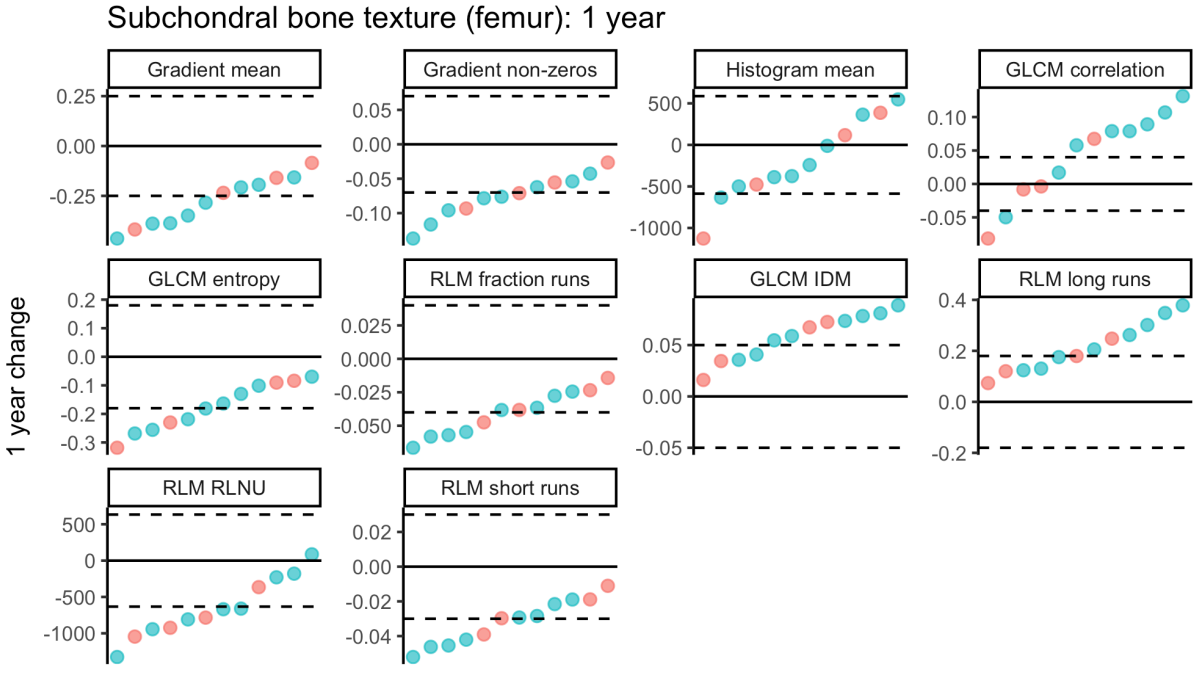
APPENDIX 7.1

6-month and 1-year changes in subchondral bone QIBs. Each dot represents an individual participant. Dotted lines are ±SDD values. **Abbreviations:** GLCM – grey-level co-occurrence matrix, RLM – run-length matrix, IDM – inverse difference moment, RLNU – run-length non-uniformity




Subchondral bone texture (tibia): 6 months
6 month change
Gradient mean
Gradient non-zeros
Histogram mean
GLCM correlation
GLCM entropy
RLM fraction runs
GLCM IDM
RLM long runs
RLM RLNU
RLM short runs
Subchondral bone texture (tibia): 1 year
1 year change
Gradient mean
Gradient non-zeros
Histogram mean
GLCM correlation
GLCM entropy
RLM fraction runs
GLCM IDM
RLM long runs
RLM RLNU
RLM short runs


Subchondral bone texture (femur): 6 month
6 month change
Gradient mean
Gradient non-zeros
Histogram mean
GLCM correlation
GLCM entropy
RLM fraction runs
GLCM IDM
RLM long runs
RLM RLNU
RLM short runs


Subchondral bone texture (femur): 1 year
1 year change
Gradient mean
Gradient non-zeros
Histogram mean
GLCM correlation
GLCM entropy
RLM fraction runs
GLCM IDM
RLM long runs
RLM RLNU
RLM short runs

CHAPTER 8:

3D CARTILAGE SURFACE MAPPING – IMPLEMENTATION IN THE AMROA STUDY

Breakdown of articular cartilage is a characteristic feature of OA, and prevention of this breakdown is an important target for potential disease modifying treatments. MR imaging methods can characterise cartilage morphology and composition, but conventional approaches are limited in their ability to detect short term changes.

Here I implement the 3D Cartilage Surface Mapping pipeline, described in chapter 3, in the AMROA study. This method is able to detect focal changes in nearly all participants at 6-months and 1-year of follow-up. By demonstrating the spatial heterogeneity and bidirectional nature of changes in cartilage morphology and composition, the method may also offer new insights into OA pathogenesis.

8.1 INTRODUCTION

Despite recent criticism of an over-emphasis on cartilage in OA research, cartilage breakdown is nevertheless a hallmark of the disease and remains very much a therapeutic target of interest(279,291,292). Therefore, the development and optimisation of quantitative imaging biomarkers (QIBs) of cartilage biology and health continues to be desirable.

Quantification of cartilage morphology (volume and thickness) using MR imaging with expert manual segmentation is well-established, and has been implemented in several clinical trials to date(293). However, it has been shown that such methods are unlikely to be sufficiently responsive over timeframes relevant to experimental medicine studies with 6-months or less follow-up (294–296). Methods for quantification of cartilage composition using MR imaging are also well-described (see chapter 2 of this thesis) and have been implemented in large observational studies with several years follow-up(297,298). However, their utility in studies with shorter follow-up periods remains uncertain.

The 3D Cartilage Surface Mapping (3D-CaSM) pipeline described and validated in chapter 3 of this thesis offers some potential advantages over existing methods as discussed previously (sections 3.1 and 3.4). However, before it can be confidently

implemented in an interventional setting, *in-vivo* evidence of its utility in an observational setting is desirable.

Therefore, the purpose of this chapter is to describe the implementation of 3D-CaSM for the analysis of cartilage morphology and composition in the AMROA study. In common with other candidate QIBs in this study, I include assessment of test-retest repeatability, ability to discriminate between OA participants and healthy controls and responsiveness to change over 6-months and 1-year.

8.2 METHODS

8.2.1 Participants & Ethical Approval

Ethical approval, participant recruitment and eligibility criteria are described in chapter 5 (sections 5.2.1 and 5.2.2).

8.2.2 Image acquisition

Details on the acquisition of MR data for the AMROA study are described in chapter 5 (section 5.2.3, Table 5.2). The high spatial resolution 3-dimensional fat suppressed spoiled gradient echo (3D SPGR) was used for assessment of cartilage morphology and for the initial steps of the 3D-CaSM pipeline. This sequence had near-identical acquisition parameters to the sequence used for validation of 3D-CaSM in chapter 3.

For assessment of cartilage composition, three complementary methods were used. Quantitative T1rho (longitudinal relaxation in the presence of a radiofrequency field) is sensitive to alterations in the proteoglycan content of cartilage, quantitative T2 (transverse relaxation) is sensitive to alterations in collagen content, collagen orientation and water content of cartilage, and delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) is sensitive to alterations in glycosaminoglycan (GAG) content. Quantitative T1rho and T2 mapping were performed using T1rho/T2 magnetisation prepared pseudo-steady-state 3D fast spin echo (FSE) sequences, performed after a period of unloading of at least 45 minutes to minimise effects related to recent weight-bearing (299). dGEMRIC was performed by administering intravenous gadolinium-based contrast agent (gdCA – Dotarem; Guerbet, Paris, France) at a dose of 0.2 mM/kg the end of the first imaging session. After leaving the scanner room, participants performed ten minutes of cycling on a stationary cycle to promote gdCA penetration into the knee joint followed by an 80-minute rest period to allow distribution of gdCA within the cartilage. Participants then returned for a second imaging session consisting of quantitative T1 (longitudinal relaxation) mapping using a variable flip angle (VFA) approach.

8.2.3 Image analysis

3D SPGR images were used for 3D-CaSM thickness measurement as described in chapter 3 (section 3.2.5).

For analysis of cartilage composition, source images for T1rho mapping, T2 mapping and dGEMRIC were rigidly registered to 3D SPGR images for motion correction using elastix(300). The femur and tibia were registered separately using masking to allow for different degrees of knee flexion between sequences, as has been described previously(110). Parameter maps were then generated for each compositional measurement by fitting the observed signal for each pixel to the appropriate equation (Table 8.1) using a linear least-squares approach. Goodness-of-fit statistics were extracted for each pixel, and pixels with poor fits ($R^2 < 0.9$) or implausible values (Table 8.1) were excluded from subsequent analysis. The inner and outer cartilage surfaces created from the 3D SPGR images were then imported into the parameter maps using the relevant rigid transform. Measurement of each parameter at each vertex on the cartilage surface was obtained by sampling the compositional data (i.e. pixel values) along the line connecting each vertex on the inner surface to its correspondent on the outer surface and taking the mean value (Figure 8.1).

Table 8.1

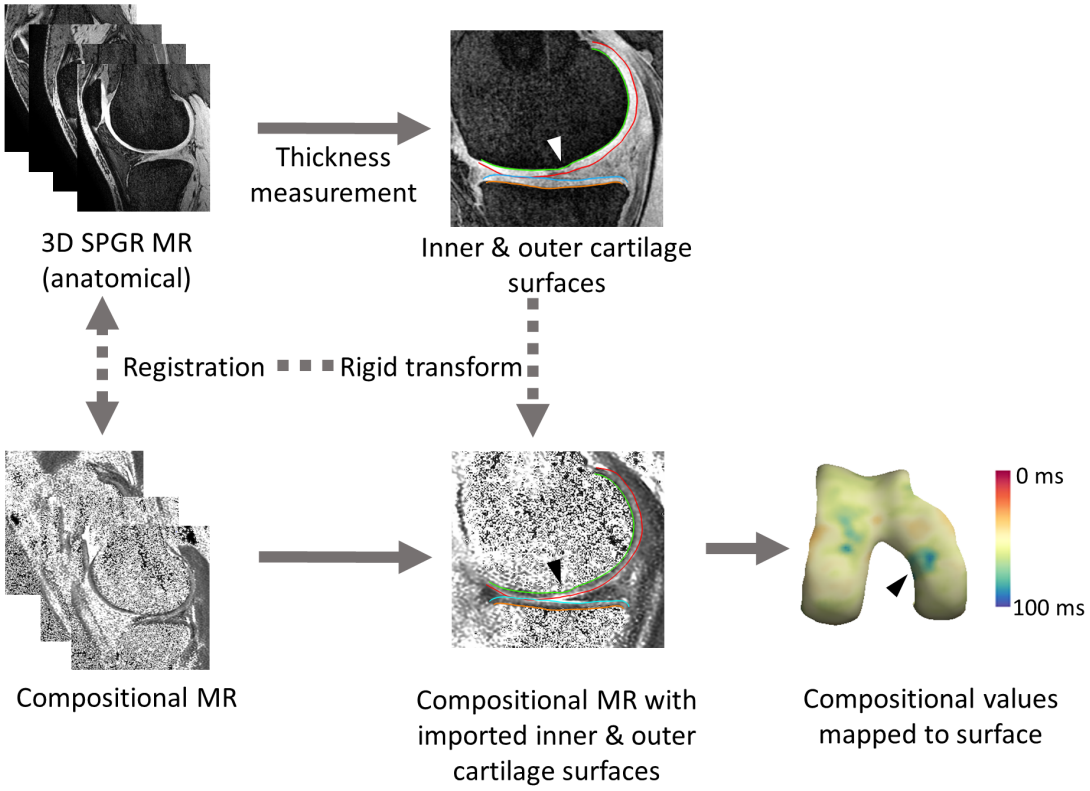
Signal equations and thresholds for implausible values for cartilage compositional methods in this study

Parameter	Signal equation	Exclusion threshold
T1rho	$S_{TSL} = A \times \exp\frac{-TSL}{T1\rho}$	> 130 ms
T2	$S_{TE} = A \times \exp\frac{-TE}{T2}$	> 100 ms
$T1_{GD}$ (dGEMRIC)	$S = M_0 \sin(\alpha)\frac{1-\varepsilon}{1-\cos(\alpha)\varepsilon}$ where $\varepsilon = \exp\frac{-TR}{T1}$	> 1000 ms*

*following correction for BMI: $(T1_{GD(corrected)} = T1_{GD(measured)} + 3(BMI - 20))$
M_0 is proportional to tissue proton density and MR system gain, and is also estimated during the fitting process.
A is a constant which is estimated during the fitting process
α is the flip angle

Figure 8.1

Outline of cartilage composition measurement process. In this example, a T1rho map is used. Note focal area of increased T1rho values on the medial femoral condyle (black arrowhead) corresponding to an area of partial thickness cartilage loss on the 3D SPGR images (white arrowhead). Femoral T1rho data used for demonstration purposes, same process used for tibial, T2 and dGEMRIC data.

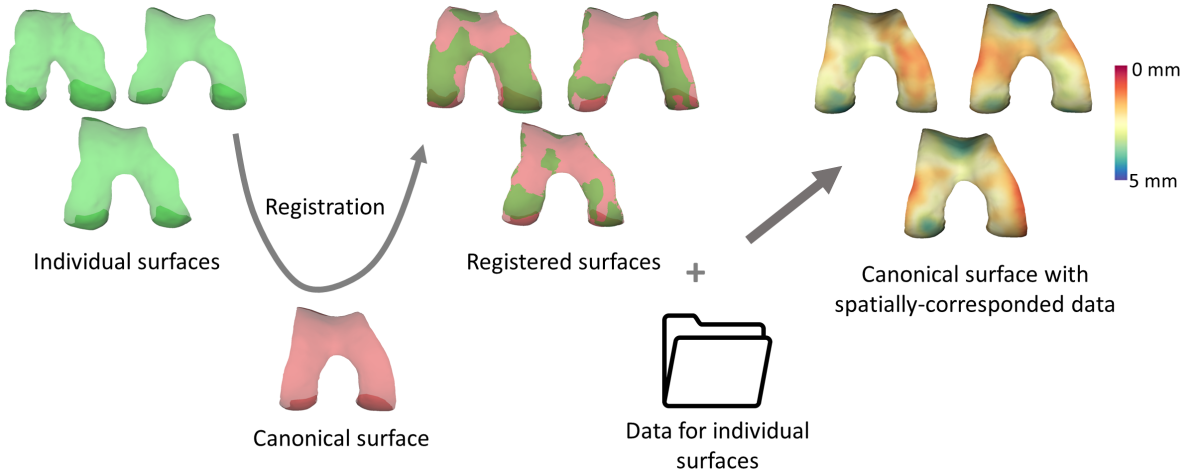


Cartilage surfaces from each timepoint from each participant were spatially normalized to a template (canonical) surface. This was carried out in wxRegSurf using a combined similarity and thin-plate-spline transformation. All surface data could then be mapped to this template surface to facilitate further spatially-corresponded analysis (Figure 8.2).

Two independent observers (JM and TT, a musculoskeletal radiologist with 10 years' experience) performed 3D-CaSM on 10 randomly selected knees for assessment of inter-observer reproducibility. Both observers also performed full manual segmentation of the femoral and tibial cartilage (time taken 3-4 hours per knee) to allow comparison of inter-observer reproducibility between 3D-CaSM and traditional manual segmentation. Manual thickness values were obtained as described for the validation study in chapter 3 (section 3.2.5).

Figure 8.2

Overview of spatial normalisation procedure. Individual cartilage surfaces are registered to a canonical surface using a combined similarity/thin-plate-spline transform. Each individual's data for each timepoint can then be 'mapped' to this canonical surface facilitating further spatially corresponded analysis. Here femur thickness data are used for demonstration – the same procedure applies for tibial and compositional data.



8.2.4 Statistical analysis

In the description of the statistical methods for this study I draw the distinction between *vertexwise* analyses where the values from all surface vertices (~5000 per participant) are used, and *surfacewise* analyses where a surface-averaged value is used.

For analysis of inter-observer reproducibility, vertexwise and surfacewise comparisons between the two observers' measurements were performed with calculation of root mean squared average coefficients of variation (RMSCVs). This was done for both 3D-CaSM and manual segmentation, and for both thickness and compositional measurements.

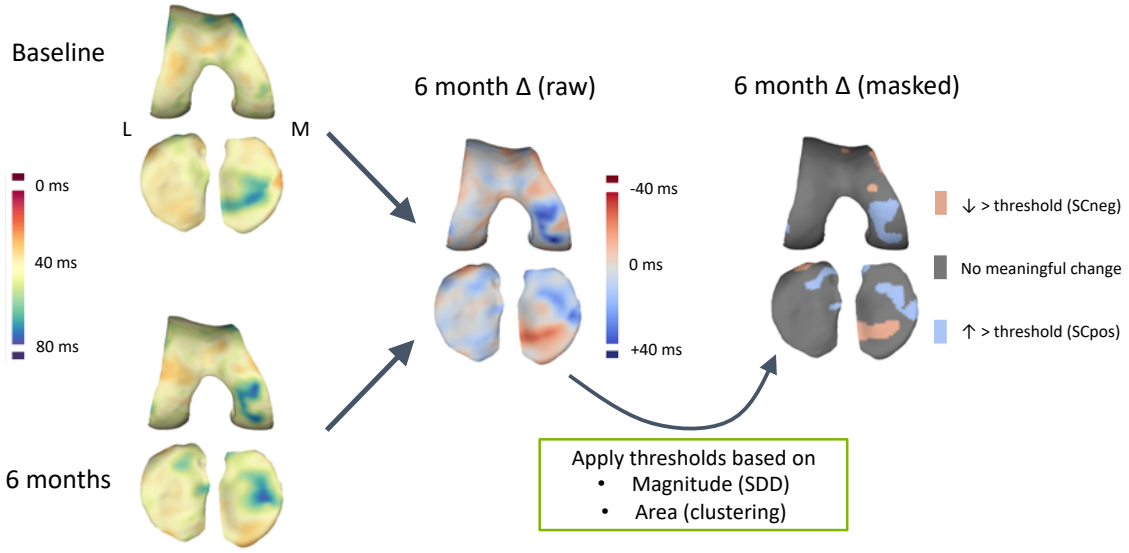
For analysis of test-retest repeatability, RMSCVs were calculated for surfacewise and vertexwise differences between baseline and one month repeat data. In addition, vertexwise and surfacewise smallest detectable differences (SDD) were calculated as described in chapter 6 (section 6.2.5).

To assess discrimination between OA and healthy knees, surfacewise data were used to estimate standardised mean differences (SMDs) between OA and healthy volunteer (HV) participants at baseline. The mean vertexwise difference between the two groups for each parameter was visualised using baseline data mapped to the canonical surfaces and subtracting the mean HV value at each vertex from the mean OA value.

To assess individual 6-month and 1-year changes, vertexwise comparison was performed between baseline and 6-month/1-year data mapped to the template surface allowing creation of an individual 'change surface' for each parameter at each timepoint (Figure 8.3). I then defined areas of significant change by applying magnitude (change greater than SDD calculated from the test-retest repeatability data for that surface/parameter combination) and area (occurring across a cluster of at least 10 vertices) thresholds to these change surfaces. The percentage of each surface with areas of significant change was calculated and used to define three summary metrics: $\%SC_{pos}$, $\%SC_{neg}$, $\%SC_{total}$, defined respectively as the percentage of the surface with areas of positive significant change, negative significant change and any significant change (regardless of sign). These metrics were compared between OA and HV groups using descriptive statistics.

Figure 8.3

Illustration of vertexwise assessment of change at 6-months. Baseline and follow-up data are mapped to the same canonical surface, enabling vertexwise comparison of data and creation of a 6-month change surface. Thresholding is then applied to this surface to highlight regions of significant change. T2 measurements used in this example; the same process was used for thickness, T1rho and dGEMRIC measurements and 1-year follow-up data.



As well as analysing surface-based changes at an individual level, the fact that all data are registered to a canonical surface permits group-averaged analysis of localised 6-month and 1-year changes and performance of statistical parametric mapping (SPM) to determine the statistical significance of these changes. SPM was performed in the OA group using SurfStat (http://www.math.mcgill.ca/keith/surfstat/), a MATLAB (MathWorks, Natick, MA, USA) toolbox for the statistical analysis of univariate and multivariate surface data using linear mixed effects models and random field theory. Linear mixed models were constructed with timepoint as a fixed effect and subject as a random effect (to account for paired data). A p value of < 0.1 was considered to be statistically significant given the exploratory nature of this analysis.

Statistical analyses other than SPM were performed using RStudio version 1.0.143. Surface-based analyses were performed using Stradwin, wxRegSurf and MATLAB version R2017a.

8.3 RESULTS

8.3.1 Participants

Participant characteristics are as described in chapter 5 (section 5.3.1). The number of analysable participants at each timepoint for each parameter are provided in Table 8.2. Compositional datasets were more commonly excluded than morphological images, with the commonest reason being unacceptable image artefact which was often due to phase wrap in the medial-lateral direction or signal drop-off inferiorly affecting quantitation of tibial cartilage. The latter links back to the issue described in chapter 7 (section 7.3.1) for quantitation of tibial subchondral bone texture, namely the compromise that had to be reached between including as much of the suprapatellar region as possible for dynamic contrast enhanced (DCE) evaluation of synovitis while avoiding positioning the tibia out of the coil sensitivity region.

Table 8.2

Number of analysable participants at each timepoint for each parameter

Parameter	Number of analysable participants at visit			
	Baseline (n = 20*)	1-month (n = 15)	6-month (n = 18**)	1-year (n = 13)
Thickness	20	15	18	11
T1rho	18	13	17	8
T2	18	12	16	8
dGEMRIC	14	14	13	9

*potential total *n* = 16 for dGEMRIC as this was not included in the protocol for all participants at this visit
**potential total *n* = 13 for dGEMRIC as this was not included in the protocol for all participants at this visit

8.3.2 Inter-observer reproducibility

Inter-observer reproducibility data are provided in Table 8.3. Surfacewise inter-observer RMSCVs were similar for measurements performed using 3D-CaSM and those performed using manual segmentation. However, vertexwise inter-observer RMSCVs were lower for 3D-CaSM for all but one measurement (medial tibial thickness) than for manual segmentation.

Table 8.3

Inter-observer reproducibility data for 3D-CaSM and conventional manual segmentation

Surface	Parameter	Inter-observer RMSCV (%)			
		Surfacewise		Vertexwise	
		3D-CaSM	Manual	3D-CaSM	Manual
Femur	Thickness	1.6	1.9	8.7	13.7
	T1rho	2.0	1.6	7.7	11.4
	T2	2.1	1.0	6.5	12.2
	dGEMRIC	3.1	1.4	9.7	12.0
Medial Tibia	Thickness	5.2	5.1	15.9	14.7
	T1rho	2.7	7.7	7.3	23.2
	T2	3.6	8.4	8.2	28.4
	dGEMRIC	1.3	3.2	4.2	24.0
Lateral Tibia	Thickness	3.9	4.5	12.9	13.4
	T1rho	5.2	1.4	10.1	14.4
	T2	4.4	2.6	9.0	15.0
	dGEMRIC	2.2	2.7	4.9	14.7

Abbreviations: RMSCV – root-mean-square average coefficient of variation

8.3.3 Test-retest repeatability

One-month test-retest 3D-CaSM repeatability data and SDD values are provided in Table 8.4. Thickness measurements had better surfacewise repeatability than compositional measurements, but similar vertexwise repeatability.

Table 8.4

One month test-retest repeatability data for 3D-CaSM.

Surface	Parameter	Surfacewise RMSCV (%)	Vertexwise RMSCV (%)	Vertexwise SDD
Femur	Thickness	3.8	11.2	0.62 mm
	T1rho	4.8	11.0	17.3 ms
	T2	4.3	9.9	12.4 ms
	dGEMRIC	8.1	10.0	152.3 ms
Medial Tibia	Thickness	5.2	14.1	0.70 mm
	T1rho	6.9	12.6	21.0 ms
	T2	6.7	11.7	14.5 ms
	dGEMRIC	7.8	7.7	124.0 ms
Lateral Tibia	Thickness	3.9	10.4	0.61 mm
	T1rho	7.7	12.7	18.2 ms
	T2	6.0	10.9	12.5 ms
	dGEMRIC	8.6	8.7	133.6 ms

Abbreviations: SDD – smallest detectable difference (repeatability coefficient from Bland-Altman analysis), otherwise as for Table 8.3

8.3.4 Baseline between-group differences

Surfacewise standardised mean differences (SMD) with 90% confidence intervals (CI) are provided for each parameter in Table 8.5. Average vertexwise differences are displayed in Figure 8.4, illustrating that between group differences are spatially heterogeneous.

Figure 8.4

Average vertexwise differences between OA and HV groups at baseline for each parameter. Red areas indicate that the parameter was on average lower in OA participants than in HV participants, whereas blue areas indicate that the parameter was on average higher in OA participants than in HV participants.

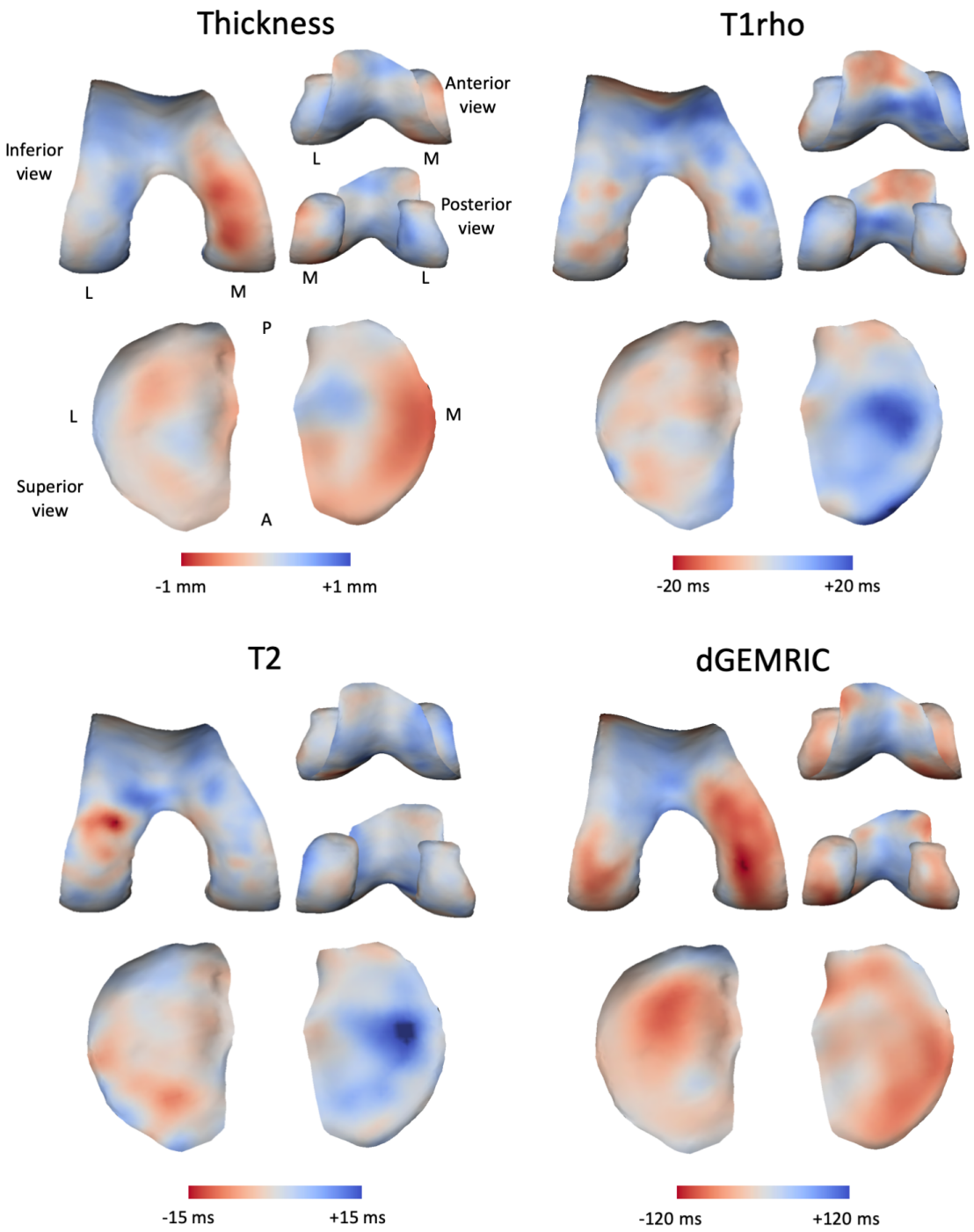


Table 8.5

Baseline surfacewise standardised mean difference values for each parameter at each surface. Positive values indicate that the parameter is higher in the OA group than the HV group.

Surface	Parameter	Surfacewise SMD (90% CI)
Femur	Thickness	0.43 (-0.38, 1.24)
	T1rho	0.48 (-0.40, 1.35)
	T2	0.45 (-0.43, 1.32)
	dGEMRIC	0.33 (-0.54, 1.21)
Medial Tibia	Thickness	-0.46 (-1.28, 0.35)
	T1rho	0.54 (-0.34, 1.41)
	T2	0.38 (-0.49, 1.25)
	dGEMRIC	0.08 (-0.86, 1.02)
Lateral Tibia	Thickness	-0.13 (-0.94, 0.67)
	T1rho	-0.08 (-0.95, 0.78)
	T2	-0.02 (-0.89, 0.84)
	dGEMRIC	0.02 (-0.91, .96)

8.3.5 Individual-level 6-month and 1-year changes

Mapping of 6-month and 1-year changes in cartilage thickness and composition revealed substantial spatial heterogeneity between participants, with areas of concurrent increase and decrease (e.g. cartilage thickening and thinning) visible in some participants (Figure 8.5). Individual %SC values for each parameter are displayed in Figure 8.6. The parameter with the highest total %SC value at 6-months and 1-year in OA participants was T2, with median (IQR) total %SC of 8.8 (5.5, 12.6) at 6-months and 9.4 (4.7, 16.7) at 1-year. Interestingly at both timepoints this total %SC value was driven mostly by positive rather than negative changes in T2.

8.3.6 Group-averaged 6-month and 1-year changes

The only statistically significant group-averaged change in the OA group was demonstrated for 6-month dGEMRIC change in the femur, where a focal area of

significant increase in dGEMRIC was demonstrated close to the medial terminal sulcus (Figure 8.7). No other statistically significant group-averaged changes were detected by SPM at 6-months or 1-year.

Figure 8.5

Baseline and 6-month follow-up thickness and compositional data for a single representative OA participant displayed on canonical femoral and tibial surfaces. Note spatial heterogeneity of changes and the co-occurrence of both significant positive and significant negative changes in thickness, T1rho and T2.

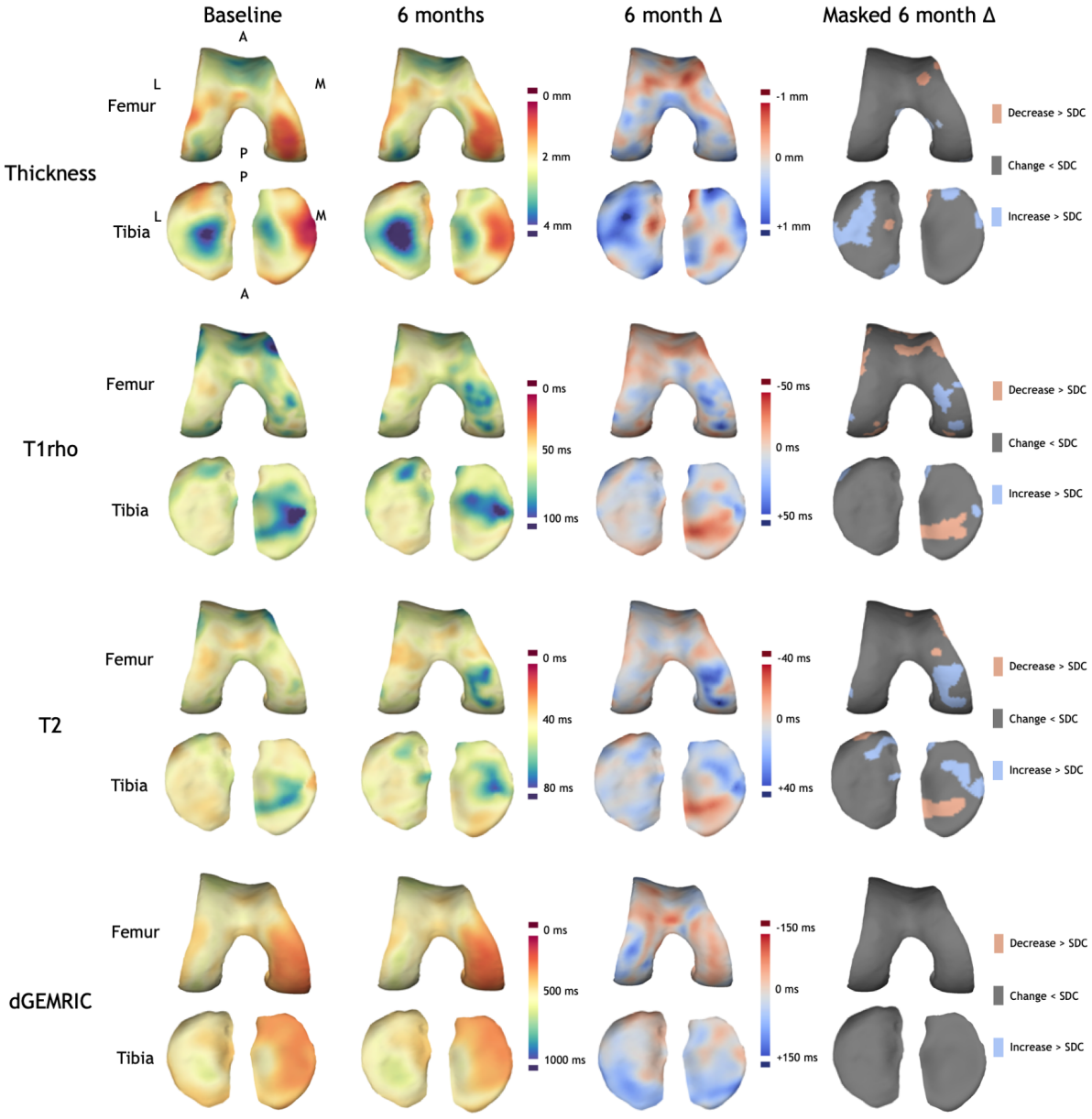


Figure 8.6

Individual 6-month and 1-year %SC data for each parameter. The dotted line represents a proposed 5% threshold above which %SC values could be considered significant within an individual. This is a relatively stringent threshold given that all areas of significant change have already passed magnitude (SDD) and area criteria.

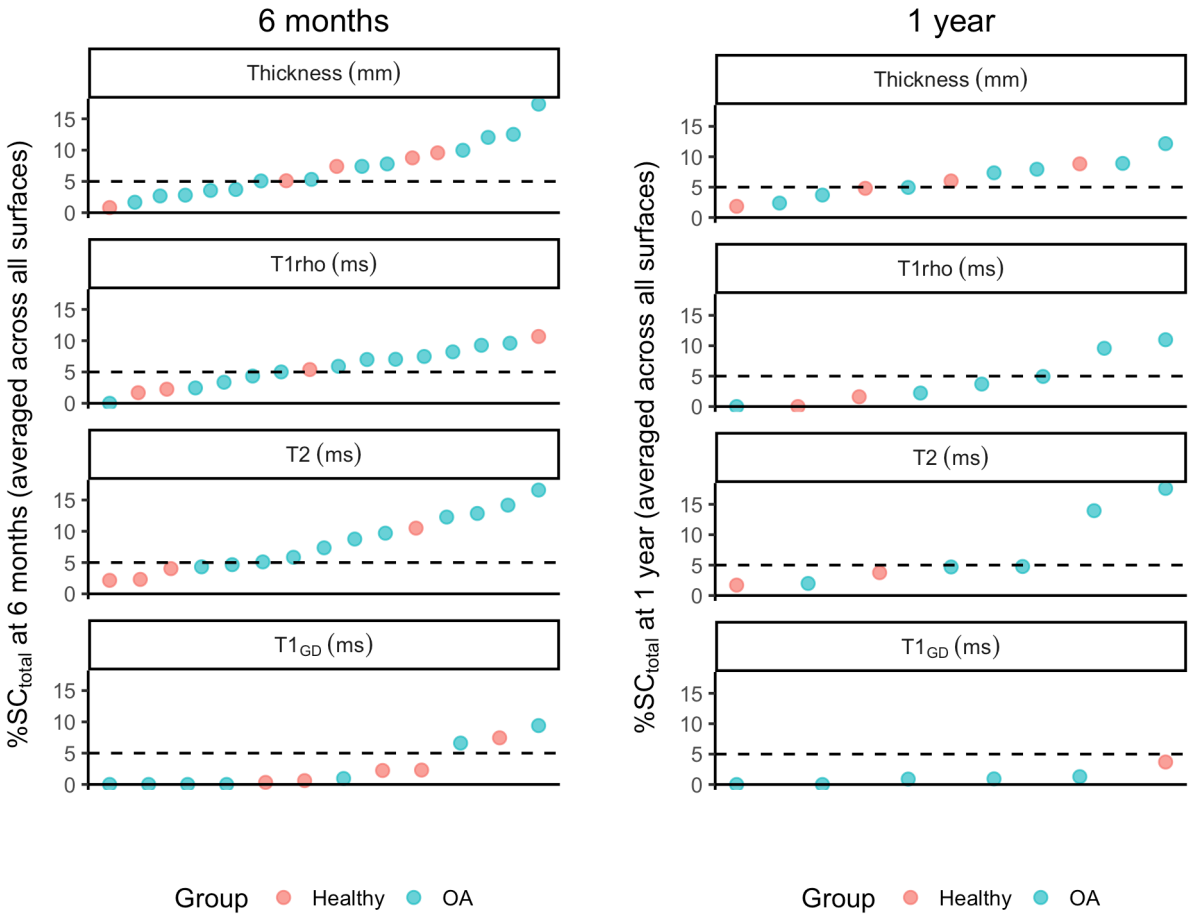
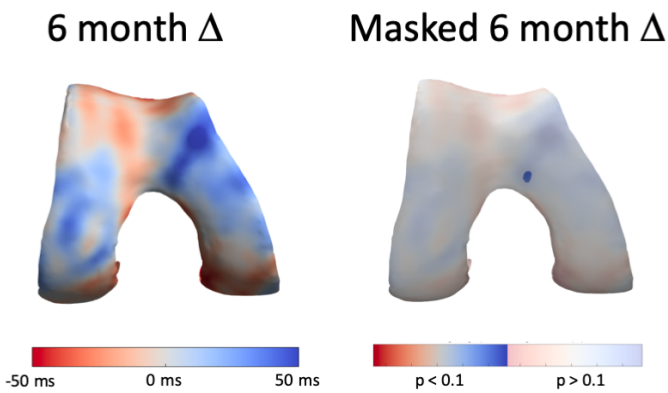


Figure 8.7

Average OA group 6-month change in femoral dGEMRIC, with focal region of statistical significance ($p < 0.1$) demonstrated close to the medial femoral sulcus. Scale for masked surface is the same as for the unmasked, but with areas without statistical significance washed out.



8.4 DISCUSSION

This chapter has described the implementation of 3D-CaSM in the AMROA study. This method of analysis allows detection of changes in cartilage morphology and composition over 6-months and 1-year which are spatially heterogeneous between individuals and often bidirectional in nature.

Inter-observer reproducibility of 3D-CaSM is similar to expert manual segmentation when values are averaged across the entire cartilage surface. However, when measurement at each individual vertex is considered, reproducibility is better for 3D-CaSM for all but one surface/parameter combination. This is unsurprising given that the automated cartilage thickness measurement algorithm uses only sparse manual contours as a guide, whereas manual thickness measurement is entirely dependent on these manual contours and therefore more subject to operator bias. Test-retest repeatability values are also commensurate with those reported for alternative methods(51,301). Previous work involving averaging of values across subregions has reported smallest detectable change (SDC) values for cartilage thickness lower than the SDD values provided here(295). There are two explanations proposed for this discrepancy. First, vertexwise SDD values are provided here which would be expected to be higher than SDD or SDC values averaged over a larger region of interest. Second, the SDC is calculated slightly differently to the SDD and will tend to be lower(302). Despite the higher SDD values reported here, the finding of 6-month changes exceeding this threshold suggests adequate sensitivity to change of 3D-CaSM nonetheless.

The potential advantage of performing vertexwise vs surfacewise analysis is illustrated by the results of baseline between-group comparisons using the two approaches (Figure 8.4 and Table 8.5). For example, average surfacewise femoral thickness measurement is counterintuitively higher in the OA than the HV group. On examination of the mapped average vertexwise differences, the reason for this becomes clear. Cartilage thickness values in the weight-bearing medial femoral condyle are lower in the OA group as would be expected given this group had been selected for having medial compartment predominant OA. However, thickness values elsewhere (lateral & patellofemoral compartments) are higher in the OA group. The phenomenon

of cartilage being simultaneously thinner in certain regions and thicker in others in OA when compared to healthy controls has been described previously, and is not always adequately reflected by ROI based approaches(303). The same is also true of compositional parameters, where cartilage regions with average values higher in the OA than the HV group and vice versa are present. These may reflect differing biomechanical loads in different areas.

This advantage of vertexwise analysis is further demonstrated when analysing 6-month change data. Areas of significant change are spatially heterogeneous between participants. Therefore, metrics such as %SC which are location agnostic may be more useful than group averaged (SPM) analyses which assume that regions of change are spatially consistent between individuals. However, this makes the implicit assumption that change in one location is as meaningful as change in any other which may not be the case. Moreover, the predictive and concurrent validity of metrics such as %SC is unknown, in contrast to conventional manual segmentation/ROI based approaches for which a large amount of literature exists(60).

Despite these potential shortcomings, the results of previous studies using similar approaches to 3D-CaSM are encouraging. Williams et al described a similar method termed ACRAC (Anatomically Corresponded Regional Analysis of Cartilage) for analysing cartilage thickness and produced individual change maps(304). Although detailed measurements were then averaged across relatively large ROIs (an approach I have deliberately sought to avoid here), the authors were nonetheless able to use this method to detect significant decreases in cartilage thickness in some regions in a cohort of 31 individuals with OA over 6-months(189). Monu et al implemented a similar pipeline (termed 'cluster analysis') in a post anterior cruciate ligament (ACL) injury population and demonstrated greater areas of increase in T1rho and T2 values in ACL injured knees compared to the healthy volunteers at 6-months and 1-year post injury(187). Although not a focus of Monu's initial analysis, analysis of the 'cluster maps' provided demonstrates similarly spatial heterogeneous and bidirectional changes to those seen in the present work. Pedoia et al described 'voxel-based relaxometry' (VBR) which performs voxelwise analysis of compositional (T1rho and T2) data. VBR was able to depict significant changes in cartilage composition which were not detected by conventional ROI based approaches(146). The location agnostic

'ordered value' approach of Eckstein et al for analysis of manually segmented cartilage data has been used in several studies to demonstrate concurrent greater thickening and greater thinning of cartilage in knees with and at-risk for OA(305). This involves ordering each subregion of cartilage analysed according to where the greatest changes have occurred, then using the region of greatest change (regardless of location) to compare between groups or associate with an outcome. The use of the %SC metric takes the 'ordered value' approach to its logical conclusion by applying a similar concept but removing the need for arbitrary subregion definition.

Data from existing surface-based and voxel-based methods are promising, and it should be noted that 3D-CaSM has the additional advantage of being able to simultaneously analyse compositional and morphological data. This may be of particular interest in interventional studies. For example, sprifermin (intra-articular recombinant human fibroblast growth factor 18) has been shown to cause average increases in cartilage thickness and volume in clinical trials(293). While these results are encouraging, the question has arisen as to whether this represents genuine cartilage growth or just cartilage swelling, which may in fact represent disease worsening(306). With conventional analysis, it has been difficult to assess the quality of this cartilage as precise localisation of the area of thickening is challenging with ROI-based approaches. However, 3D-CaSM could delineate in each individual knee exactly where any cartilage thickening had occurred. Because the thickness and compositional data are registered to the same surface, assessment of compositional values in the precise region of cartilage thickening could then be performed. This may allow some assessment of the quality of the cartilage in areas of thickening and differentiation of cartilage regeneration vs swelling (120).

The previously limited use of analytical approaches taking into account bidirectional changes in cartilage provides a speculative explanation as to why previous studies of short-term changes in cartilage morphology have failed to detect significant unidirectional changes, and clinical trials using only cartilage loss in a single subregion as an endpoint have failed to demonstrate a significant effect(65,296). Moreover, it supports the paradigm of OA as a disorder characterised by abnormal cartilage turnover rather than simply a disorder of cartilage loss, for which there is considerable biological and imaging evidence(307,308). For example, in early osteoarthritis the

synthesis of proteoglycans and type II collagen actually increases (309,310). In addition, disruption of the normal cartilage structure may cause a counterintuitive increase in the amount of bound water molecules by increasing the number of accessible hydrophilic binding sites(311). Therefore, when considering changes in quantitative imaging biomarkers reflective of these tissue components, one would expect to see changes reflective of this (e.g. decreases in T1rho, T2, increases in dGEMRIC) in some patients rather than only changes in the traditionally expected direction. In order to investigate this further, future work should explore the clinical meaning of MR imaging-defined cartilage turnover in terms of association with patient symptoms and other measures of structural deterioration.

The main limitation of this study is the lack of longitudinal validation of 3D-CaSM and the metrics derived from it (e.g. %SC). While the method may offer improved responsiveness to conventional ROI-based methods, it should be borne in mind that changes detected by these conventional methods have been linked to longer term clinical and radiological outcomes whereas changes detected by 3D-CaSM have not. Implementation of 3D-CaSM in larger cohorts with longer term follow-up is planned to help establish the clinical meaning of metrics such as %SC. Another limitation is the lack of histological or biological validation of the findings. Therefore, the suggestion made in preceding paragraphs that 3D-CaSM may be able to depict abnormal 'cartilage turnover' should be regarded as somewhat speculative. Finally, SPM was used to assess the significance of group-averaged changes over 6-months and 1-year. While this approach is valid and has been used extensively in the neuroimaging community, larger group sizes are usual. Power calculations for SPM are difficult, but it is likely that this study was underpowered in its ability to detect group-averaged changes. However, as mentioned above it should also be noted that SPM may not be the optimal analysis method for 3D-CaSM data given the spatial heterogeneity of the changes demonstrated. Location agnostic measures such as %SC may provide better responsiveness.

In conclusion, this chapter has described the implementation of 3D-CaSM in the AMROA study. This analysis method is able to detect and map significant changes in cartilage morphology and composition at 6-months and demonstrates test-retest

repeatability and inter-observer reproducibility comparable to existing gold-standard methods.

CHAPTER 9:

CONCLUSIONS

9.1 CONTRIBUTIONS TO KNOWLEDGE

This thesis has investigated the role of quantitative imaging biomarkers (QIBs) in the assessment of knee OA with a focus on their role in experimental medicine (EM) studies. The major contributions to knowledge of this thesis include an extension of the evidence base behind several existing MR QIBs, the development of a novel analysis pipeline (3D-CaSM) for analysis of articular cartilage morphology and composition and an assessment of the performance characteristics of candidate MR QIBs for EM studies.

The systematic review and meta-analysis of the reliability and discriminative validity of cartilage compositional MR imaging (chapter 2) has been acknowledged by the Radiological Society of North America (RSNA) Quantitative Imaging Biomarkers Alliance (QIBA) as an important contribution to the validation of these QIBs(301,312). The resultant publication will form part of an upcoming QIBA profile for cartilage compositional MR imaging. QIBA profiles are developed using published data to generate evidence-based performance claims on the magnitude of change required for the biomarker within an individual for that change to be considered significant. For example, the QIBA profile for tumour volume change with computed tomography (CT) for advanced malignancy states that "A true change in a tumor volume has occurred with 95% confidence if the measured change is larger than 24%, 29% or 39% when the longest in-plane diameter is initially 50-100 mm, 35-49 mm or 10-34 mm, respectively"(313). Detailed knowledge of the repeatability and reproducibility of QIBs, as provided by chapter 2 of this thesis, is a key piece of information used to derive such claims.

The development and validation of 3D cartilage surface mapping (3D-CaSM) has allowed the detection and localisation of focal changes in cartilage thickness and composition over as little as 6-months in the AMROA study. 3D-CaSM has demonstrated that changes in cartilage in OA participants are spatially heterogeneous and are bidirectional in nature. This work will contribute to a growing literature challenging the concept of OA as characterised by only loss of cartilage over time. The analysis pipeline will next be applied to the Foundation for the National Institutes of Health (FNIH) Biomarkers Consortium PROGRESS-OA cohort to establish the

predictive and concurrent validity of 3D-CaSM for clinically meaningful outcomes(314).

The predictive and concurrent validity of subchondral bone texture for knee OA progression were established in chapter 4 in the Osteoarthritis Initiative Bone Ancillary Study cohort. Previous work had demonstrated the utility of subchondral bone texture analysis in a cross-sectional setting and had established construct validity. The work presented in this thesis is an important step in promoting the use of this QIB in longitudinal and interventional studies(284).

The AMROA study has established the important performance characteristics of candidate QIBs for bone, synovium and cartilage for EM studies. The results enable head-to-head comparison of the QIBs studied across the domains of test-retest repeatability, discriminative validity and responsiveness to change at both 6-months and 1-year (Table 9.1). While such comparison can be informative, it should be noted that choice of QIB for interventional studies will ultimately depend most on the anticipated mechanism of action of the treatment being tested.

Table 9.1

Comparison of QIBs included in AMROA study across domains of interest. For each domain, the QIB from each target tissue (bone, cartilage, synovium) with the best performance is selected for brevity.

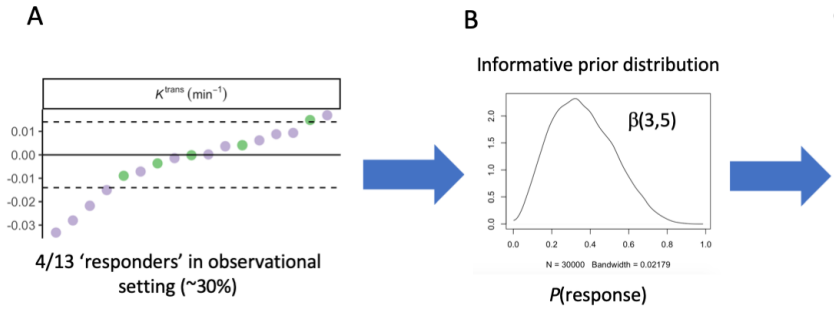
Target tissue	Test-retest repeatability (RMSCV, %)		Discrimination OA vs HV (SMD)		Responsiveness to change at 6-months (% OA participants with Δ > SDD)	
Synovium	K^{trans}	17.8	K^{trans}	0.94	K^{trans}	42
Bone	RLM SRLE (tibia)	0.8	GLCM correlation (tibia)	1.26	RLM SRLE (tibia)[#]	45
Cartilage	Thickness	4.3*	T1rho	0.54*	T2	81**

*surfacewise analysis **vertexwise analysis
[#]tied with Tibial RLM LRLE & Femoral GLCM inverse difference moment
Abbreviations: RMSCV – root mean square average coefficient of variation, SMD – standardised mean difference, SDD – smallest detectable difference, RLM – run-length-matrix, GLCM – grey-level-co-occurrence matrix, SRLE – short run length emphasis, LRLE – long run length emphasis

The findings from the AMROA study can also inform sample size calculation for future studies. In particular, the within-subject standard deviation estimated from test-retest data is an important parameter for such calculations. An example is provided in Chapter 6 (Section 6.4). However, such sample size calculations assumes a frequentist approach which is rarely used in the analysis of EM studies where Bayesian approaches are preferred(315). The reason for preferring Bayesian approaches is that the goal of EM studies is not necessarily to establish definitive evidence of effect, rather to establish the level of confidence in any efficacy signal detected. The data from the AMROA study can also be helpful in this regard. Taking again the example of synovial K^{trans}, the number of OA participants demonstrating significant reductions in this parameter at 6-months in this observational study (4/13 or 30%) can be used to create an informative prior distribution of predicted 'response' in the absence of any treatment effect. This information can then be used to determine the probability that a response rate observed in an interventional study was due to treatment effect using Markov-Chain-Monte-Carlo sampling from the prior distribution (Figure 9.1).

Figure 9.1

Use of AMROA data to inform Bayesian analysis of EM study data. Given number of participants showing a reduction in K^{trans} in the observational setting (a), an informative prior distribution (b) can be derived to summarise the probability of response in the absence of disease modifying treatment. This can then be used (c) to assess the likelihood of a response rate observed in a small single-arm EM study (n = 15 used as an example here) of a putative anti-inflammatory DMOAD being due to treatment effect.



C

Putative anti-inflammatory DMOAD study, n = 15

$n_{responders}$	P(treatment effect)
15	99.92%
13	99.61%
11	97.2%
9	89.7%
7	75.1%
5	54.4%

The limitations of individual components of the AMROA study are discussed in chapters 5-8. The major overarching limitation is the small sample size. As discussed previously, this was to some extent intentional. However, the small sample size does still reduce the precision of estimates of QIB performance characteristics derived from the study data. Nevertheless, the AMROA study has still provided important data to

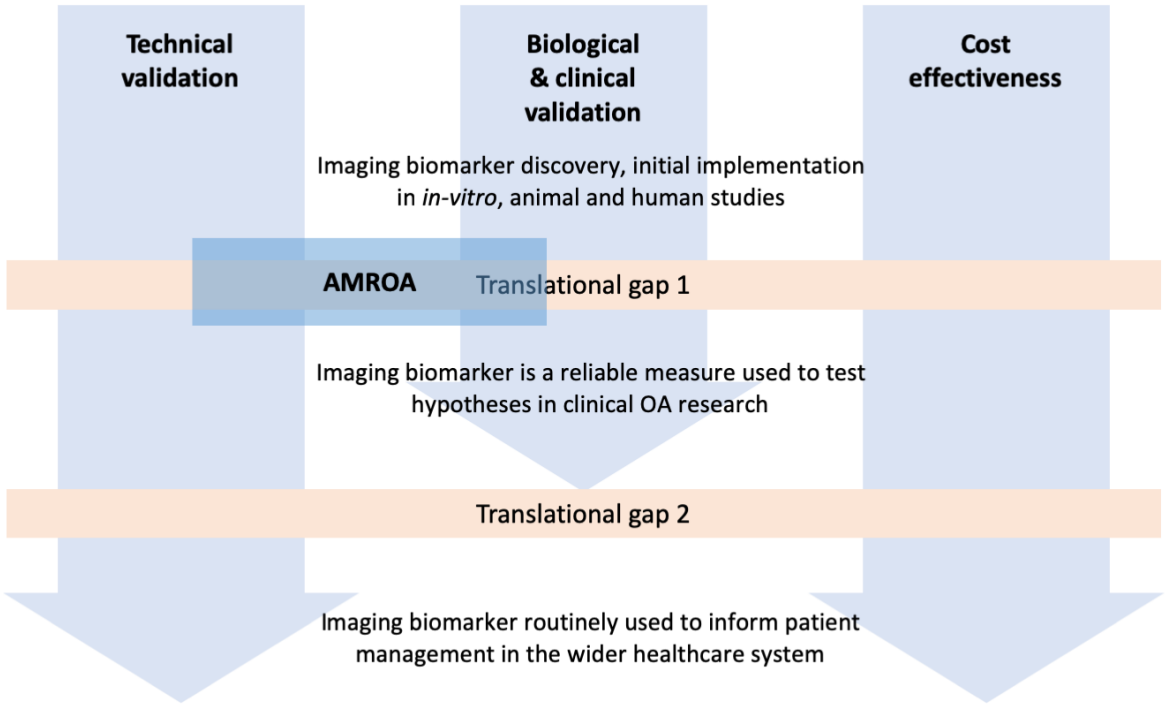
enable the use of QIBs in interventional EM studies where similar uncertainty is likely to exist.

9.2 FUTURE DIRECTIONS

A recent imaging biomarker roadmap published by oncology researchers outlines two "translational gaps" in the development of QIBs(81). This study represents an effort to cross translational gap 1 (Figure 9.2). The next step is to achieve incorporation of a selection of the QIBs into ongoing interventional studies.

Figure 9.2

Translational gaps in imaging biomarker development. The AMROA study results will help enable the QIBs evaluated to cross translational gap 1 by providing technical and clinical validation. Adapted from (81).



The QIBs studied in this thesis also have the potential to enter more widespread clinical use. However, this is a longer-term aim and will require several additional studies, including the demonstration of multicentre feasibility and robustness and cost efficacy. As with all OA imaging, the widespread uptake of the QIBs will be limited until a treatment which demonstrates disease modifying capability is available and advanced imaging has a demonstrable influence on patient management, whether for assessing eligibility for or monitoring response to treatment.

9.3 CONCLUSIONS

1. Cartilage compositional MR methods are reliable and, in the case of T1rho and T2 mapping, are able to discriminate individuals with knee OA and controls.
2. 3D Cartilage Surface Mapping (3D-CaSM) is as accurate as manual segmentation for measurement of cartilage thickness when compared to the reference standard and also reduces time taken for analysis with better inter-observer reproducibility.
3. Baseline and 1-year change in subchondral bone texture are associated with radiographic knee OA progression at 3 years, with modest predictive ability.
4. For dynamic contrast enhanced (DCE) MR assessment of synovitis, the QIB with the best performance characteristics for experimental medicine studies is K^{trans}, suggesting that full pharmacokinetic modelling of DCE data is worth the additional effort.
5. Subchondral bone area and subchondral bone texture demonstrate significant changes in some individuals over 6-months. Subchondral bone texture features are more responsive than subchondral bone area, but are less interpretable.
6. 3D-CaSM can detect and localise significant changes in cartilage morphology and composition over 6-months in some individuals.
   Changes in cartilage morphology and composition in OA are spatially heterogeneous between participants and bidirectional in nature. Location agnostic analytical approaches may therefore have some advantages.

REFERENCES

1. Arthritis Research UK. Osteoarthritis in general practice. Arthritis Research UK; 2013.

2. Department for Work and Pensions. Tabulation Tool [Internet]. [cited 2016 Feb 7]. Available from: http://tabulation-tool.dwp.gov.uk/100pc/tabtool.html

3. Pre-Competitive Consortium for Osteoarthritis. Osteoarthritis as a Serious Disease. Osteoarthritis Research Society International; 2018.

4. Castañeda S, Roman-Blas JA, Largo R, Herrero-Beaumont G. Osteoarthritis: a progressive disease with changing phenotypes. Rheumatology. 2014 Jan;53(1):1–3.

5. Hunter DJ, Felson DT. Osteoarthritis. BMJ. 2006 Mar 18;332(7542):639–42.

6. Skou ST, Roos EM, Laursen MB, Rathleff MS, Arendt-Nielsen L, Simonsen O, et al. A Randomized, Controlled Trial of Total Knee Replacement. New England Journal of Medicine. 2015 Oct 22;373(17):1597–606.

7. Carr AJ, Robertsson O, Graves S, Price AJ, Arden NK, Judge A, et al. Knee replacement. The Lancet. 2012 Apr 7;379(9823):1331–40.

8. Siemieniuk RAC, Harris IA, Agoritsas T, Poolman RW, Brignardello-Petersen R, Velde SV de, et al. Arthroscopic surgery for degenerative knee arthritis and meniscal tears: a clinical practice guideline. BMJ. 2017 May 10;357:j1982.

9. Elattrache N, Lattermann C, Hannon M, Cole B. New England journal of medicine article evaluating the usefulness of meniscectomy is flawed. Arthroscopy. 2014;30(5):542–543.

10. Lubowitz JH, D'Agostino RB, Provencher MT, Rossi MJ, Brand JC. Can We Trust Knee Meniscus Studies? One-Way Crossover Confounds Intent-to-Treat Statistical Methods. Arthroscopy. 2016;32(11):2187–2190.

11. Abram SGF, Beard DJ, Price AJ. Arthroscopic meniscal surgery. The Bone & Joint Journal. 2019 Jun 1;101-B(6):652–9.

12. Watt FE, Gulati M. New Drug Treatments for Osteoarthritis: What is on the Horizon? Eur Med J Rheumatol. 2017 Mar 2;2(1):50–8.

13. Loeser RF, Goldring SR, Scanzello CR, Goldring MB. Osteoarthritis: A Disease of the Joint as an Organ. Arthritis Rheum. 2012 Jun;64(6):1697–707.

14. Hunter DJ. Osteoarthritis: time for us all to shift the needle. Rheumatology (Oxford). 2018 May 1;57(suppl_4):iv1–2.

15. Martel-Pelletier J, Barr AJ, Cicuttini FM, Conaghan PG, Cooper C, Goldring MB, et al. Osteoarthritis. Nature Reviews Disease Primers. 2016 Oct 13;2:16072.

16. Heinegård D, Saxne T. The role of the cartilage matrix in osteoarthritis. Nat Rev Rheumatol. 2011 Jan;7(1):50–6.

17. Palmer AJR, Brown CP, McNally EG, Price AJ, Tracey I, Jezzard P, et al. Non-invasive imaging of cartilage in early osteoarthritis. Bone Joint J. 2013 Jun;95-B(6):738–46.

18. Burr DB, Gallant MA. Bone remodelling in osteoarthritis. Nat Rev Rheumatol. 2012 Nov;8(11):665–73.

19. Teichtahl AJ, Wluka AE, Wijethilake P, Wang Y, Ghasem-Zadeh A, Cicuttini FM. Wolff's law in action: a mechanism for early knee osteoarthritis. Arthritis Res Ther. 2015 Sep 1;17:207.

20. Burr DB. Anatomy and physiology of the mineralized tissues: role in the pathogenesis of osteoarthrosis. Osteoarthr Cartil. 2004;12 Suppl A:S20-30.

21. Suri S, Walsh DA. Osteochondral alterations in osteoarthritis. Bone. 2012 Aug;51(2):204–11.

22. Wang T, Wen C-Y, Yan C-H, Lu W-W, Chiu K-Y. Spatial and temporal changes of subchondral bone proceed to microscopic articular cartilage degeneration in guinea pigs with spontaneous osteoarthritis. Osteoarthr Cartil. 2013 Apr;21(4):574–81.

23. Sellam J, Berenbaum F. The role of synovitis in pathophysiology and clinical symptoms of osteoarthritis. Nat Rev Rheumatol. 2010 Nov;6(11):625–35.

24. Mathiessen A, Conaghan PG. Synovitis in osteoarthritis: current understanding with therapeutic implications. Arthritis Res Ther. 2017 Feb 2;19(1):18.

25. Blom AB, van Lent PL, Libregts S, Holthuysen AE, van der Kraan PM, van Rooijen N, et al. Crucial role of macrophages in matrix metalloproteinase-mediated cartilage destruction during experimental osteoarthritis: involvement of matrix metalloproteinase 3. Arthritis Rheum. 2007 Jan;56(1):147–57.

26. van Lent PL, Blom AB, van der Kraan P, Holthuysen AEM, Vitters E, van Rooijen N, et al. Crucial role of synovial lining macrophages in the promotion of transforming growth factor beta-mediated osteophyte formation. Arthritis Rheum. 2004 Jan;50(1):103–11.

27. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis. 1957 Dec;16(4):494–502.

28. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthr Cartil. 2007 Jan 1;15:A1–56.

29. Gossec L, Jordan JM, Mazzuca SA, Lam M-A, Suarez-Almazor ME, Renner JB, et al. Comparative evaluation of three semi-quantitative radiographic grading techniques for knee osteoarthritis in terms of validity and reproducibility in 1759 X-rays: report of the OARSI-OMERACT task force. Osteoarthr Cartil. 2008 Jul;16(7):742–8.

30. Schiphof D, de Klerk BM, Kerkhof HJM, Hofman A, Koes BW, Boers M, et al. Impact of different descriptions of the Kellgren and Lawrence classification criteria on the diagnosis of knee osteoarthritis. Ann Rheum Dis. 2011 Aug;70(8):1422–7.

31. Duryea J, Li J, Peterfy CG, Gordon C, Genant HK. Trainable rule-based algorithm for the measurement of joint space width in digital radiographic images of the knee. Med Phys. 2000 Mar;27(3):580–91.

32. Conaghan PG, Hunter DJ, Maillefert JF, Reichmann WM, Losina E. Summary and recommendations of the OARSI FDA osteoarthritis Assessment of Structural Change Working Group. Osteoarthr Cartil. 2011 May;19(5):606–10.

33. Neumann G, Hunter D, Nevitt M, et al. Location specific radiographic joint space width for osteoarthritis progression. Osteoarthr Cartil. 2009;17(6):761–765.

34. Eckstein F, Boudreau R, Wang Z, Hannon MJ, Duryea J, Wirth W, et al. Comparison of radiographic joint space width and magnetic resonance imaging for prediction of knee replacement: A longitudinal case-control study from the Osteoarthritis Initiative. Eur Radiol. 2016 Jun 1;26(6):1942–51.

35. Charles HC, Kraus VB, Ainslie M, Hellio Le Graverand-Gastineau M-P. Optimization of the fixed-flexion knee radiograph. Osteoarthr Cartil. 2007;15(11):1221–1224.

36. Roth M, Wirth W, Emmanuel K, Culvenor AG, Eckstein F. The contribution of 3D quantitative meniscal and cartilage measures to variation in normal radiographic joint space width-Data from the Osteoarthritis Initiative healthy reference cohort. Eur J Radiol. 2017 Feb;87:90–8.

37. Hunter DJ, Zhang YQ, Tu X, Lavalley M, Niu JB, Amin S, et al. Change in joint space width: hyaline articular cartilage loss or alteration in meniscus? Arthritis Rheum. 2006 Aug;54(8):2488–95.

38. Wirth W, Duryea J, Hellio Le Graverand M-P, John MR, Nevitt M, Buck RJ, et al. Direct comparison of fixed flexion, radiography and MRI in knee osteoarthritis: responsiveness data from the Osteoarthritis Initiative. Osteoarthr Cartil. 2013 Jan;21(1):117–25.

39. Buckland-Wright JC, Lynch JA, Macfarlane DG. Fractal signature analysis measures cancellous bone organisation in macroradiographs of patients with knee osteoarthritis. Ann Rheum Dis. 1996 Oct;55(10):749–55.

40. Kraus VB, Collins JE, Charles HC, Pieper CF, Whitley L, Losina E, et al. Predictive Validity of Radiographic Trabecular Bone Texture in Knee Osteoarthritis: The Osteoarthritis Research Society International/Foundation for the National Institutes of Health Osteoarthritis Biomarkers Consortium. Arthritis Rheum. 2018;70(1):80–7.

41. Lynch JA, Hawkes DJ, Buckland-Wright JC. A robust and accurate method for calculating the fractal signature of texture in macroradiographs of osteoarthritic knees. Med Inform (Lond). 1991 Jun;16(2):241–51.

42. Guermazi A, Niu J, Hayashi D, Roemer FW, Englund M, Neogi T, et al. Prevalence of abnormalities in knees detected by MRI in adults without knee osteoarthritis: population based observational study (Framingham Osteoarthritis Study). BMJ. 2012 Aug 29;345:e5339.

43. National Institute for Health and Care Excellence. Osteoarthritis: care and management (CG177). National Institute for Health and Care Excellence; 2014.

44. Hunter DJ, Altman RD, Cicuttini F, Crema MD, Duryea J, Eckstein F, et al. OARSI Clinical Trials Recommendations: Knee imaging in clinical trials in osteoarthritis. Osteoarthr Cartil. 2015 May;23(5):698–715.

45. Guermazi A, Roemer FW, Haugen IK, Crema MD, Hayashi D. MRI-based semiquantitative scoring of joint pathology in osteoarthritis. Nat Rev Rheumatol. 2013 Apr;9(4):236–51.

46. Hunter DJ, Guermazi A, Lo GH, Grainger AJ, Conaghan PG, Boudreau RM, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). Osteoarthr Cartil. 2011 Aug;19(8):990–1002.

47. Peterfy CG, Guermazi A, Zaim S, Tirman PFJ, Miaux Y, White D, et al. Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. Osteoarthr Cartil. 2004 Mar;12(3):177–90.

48. Hunter DJ, Lo GH, Gale D, Grainger AJ, Guermazi A, Conaghan PG. The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston Leeds Osteoarthritis Knee Score). Ann Rheum Dis. 2008 Feb;67(2):206–11.

49. Hunter DJ, Zhang W, Conaghan PG, Hirko K, Menashe L, Li L, et al. Systematic review of the concurrent and predictive validity of MRI biomarkers in OA. Osteoarthr Cartil. 2011 May;19(5):557–88.

50. Bedson J, Croft PR. The discordance between clinical and radiographic knee osteoarthritis: A systematic search and summary of the literature. BMC Musculoskelet Disord. 2008 Sep 2;9:116.

51. Hunter DJ, Zhang W, Conaghan PG, Hirko K, Menashe L, Reichmann WM, et al. Responsiveness and reliability of MRI in knee osteoarthritis: a meta-analysis of published evidence. Osteoarthr Cartil. 2011 May;19(5):589–605.

52. Collins JE, Losina E, Nevitt MC, Roemer FW, Guermazi A, Lynch JA, et al. Semiquantitative Imaging Biomarkers of Knee Osteoarthritis Progression: Data From the Foundation for the National Institutes of Health Osteoarthritis Biomarkers Consortium. Arthritis Rheum. 2016 Oct;68(10):2422–31.

53. Reichmann WM, Maillefert JF, Hunter DJ, Katz JN, Conaghan PG, Losina E. Responsiveness to change and reliability of measurement of radiographic joint space width in osteoarthritis of the knee: a systematic review. Osteoarthr Cartil. 2011 May;19(5):550–6.

54. Roemer FW, Aydemir A, Lohmander S, Crema MD, Marra MD, Muurahainen N, et al. Structural effects of sprifermin in knee osteoarthritis: a post-hoc analysis on cartilage and non-cartilaginous tissue alterations in a randomized controlled trial. BMC Musculoskelet Disord. 2016 Jul 9;17:267.

55. Peterfy CG, Schneider E, Nevitt M. The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. Osteoarthr Cartil. 2008 Dec;16(12):1433–41.

56. Stahl R, Jain SK, Lutz J, Wyman BT, Le Graverand-Gastineau M-PH, Vignon E, et al. Osteoarthritis of the knee at 3.0 T: comparison of a quantitative and a semi-quantitative score for the assessment of the extent of cartilage lesion and bone marrow edema pattern in a 24-month longitudinal study. Skeletal Radiol. 2011 Oct;40(10):1315–27.

57. Collins JE, Hunter DJ, Guermazi A, Roemer FW, Katz JN, Losina E. When the 'Whole' is more than a 'Sum' of its parts: quantifying total burden of knee oa using semi-quantitative knee MRI assessment. Osteoarthr Cartil. 2018 Apr 1;26:S50–1.

58. Buck RJ, Wirth W, Dreher D, Nevitt M, Eckstein F. Frequency and spatial distribution of cartilage thickness change in knee osteoarthritis and its relation to clinical and radiographic covariates - data from the osteoarthritis initiative. Osteoarthr Cartil. 2013 Jan;21(1):102–9.

59. Deveza LA, Downie A, Tamez-Peña JG, Eckstein F, Van Spil WE, Hunter DJ. Trajectories of femorotibial cartilage thickness among persons with or at risk of

knee osteoarthritis: development of a prediction model to identify progressors. Osteoarthr Cartil. 2019 Feb;27(2):257–65.

60. Eckstein F, Collins JE, Nevitt MC, Lynch JA, Kraus VB, Katz JN, et al. Brief Report: Cartilage Thickness Change as an Imaging Biomarker of Knee Osteoarthritis Progression: Data From the Foundation for the National Institutes of Health Osteoarthritis Biomarkers Consortium. Arthritis Rheum. 2015 Dec;67(12):3184–9.

61. Eckstein F, Wirth W, Lohmander LS, Hudelmaier MI, Frobell RB. Five-year followup of knee joint cartilage thickness changes after acute rupture of the anterior cruciate ligament. Arthritis Rheum. 2015 Jan;67(1):152–61.

62. Haverkamp DJ, Schiphof D, Bierma-Zeinstra SM, Weinans H, Waarsing JH. Variation in joint shape of osteoarthritic knees. Arthritis Rheum. 2011;63(11):3401–7.

63. Bowes MA, Vincent GR, Wolstenholme CB, Conaghan PG. A novel method for bone area measurement provides new insights into osteoarthritis and its progression. Ann Rheum Dis. 2015 Mar;74(3):519–25.

64. Bowes MA, Maciewicz RA, Waterton JC, Hunter DJ, Conaghan PG. Bone Area Provides a Responsive Outcome Measure for Bone Changes in Short-term Knee Osteoarthritis Studies. J Rheumatol. 2016 Dec;43(12):2179–82.

65. Lohmander LS, Hellot S, Dreher D, Krantz EFW, Kruger DS, Guermazi A, et al. Intraarticular sprifermin (recombinant human fibroblast growth factor 18) in knee osteoarthritis: a randomized, double-blind, placebo-controlled trial. Arthritis Rheum. 2014 Jul;66(7):1820–31.

66. Conaghan PG, Bowes MA, Kingsbury SR, Brett A, Guillard G, Jansson Å, et al. 14L: MIV-711, a Novel Cathepsin K Inhibitor Demonstrates Evidence of Osteoarthritis Structure Modification: Results from a 6 Month Randomised Double-Blind Placebo-Controlled Phase IIa Trial. Arthritis Rheum. 2017;69(S10).

67. Kwoh CK. Epidemiology of Osteoarthritis. In: Newman AB, Cauley JA, editors. Epidemiol Aging. Dordrecht: Springer Netherlands; 2012. p. 523–536

68. Burstein D, Mitchell R, Gray ML. Strategically Guiding Research through Careful Consideration of the Path to Impact. Radiology. 2018 Nov 20;290(1):5–7.

69. O'Connor JPB, Jackson A, Parker GJM, Roberts C, Jayson GC. Dynamic contrast-enhanced MRI in clinical trials of antivascular therapies. Nature Reviews Clinical Oncology. 2012 Feb 14;9(3):167–77.

70. Budzik J-F, Ding J, Norberciak L, Pascart T, Toumi H, Verclytte S, et al. Perfusion of subchondral bone marrow in knee osteoarthritis: A dynamic contrast-enhanced magnetic resonance imaging preliminary study. Eur J Radiol. 2017 Mar;88:129–34.

71. Gait AD, Hodgson R, Parkes MJ, Hutchinson CE, O'Neill TW, Maricar N, et al. Synovial volume vs synovial measurements from dynamic contrast enhanced MRI as measures of response in osteoarthritis. Osteoarthr Cartil. 2016 Aug;24(8):1392–8.

72. Wenham CYJ, Balamoody S, Grainger AJ, Hensor EMA, Draycott S, Hodgson R, et al. The responsiveness of novel, dynamic, contrast-enhanced magnetic resonance measures of total knee synovitis after intra-articular corticosteroid for painful osteoarthritis. Osteoarthr Cartil. 2014 Oct;22(10):1614–8.

73. Leach MO, Brindle KM, Evelhoch JL, Griffiths JR, Horsman MR, Jackson A, et al. The assessment of antiangiogenic and antivascular therapies in early-stage clinical trials using magnetic resonance imaging: issues and recommendations. British Journal of Cancer. 2005 May;92(9):1599–610.

74. Medical Research Council MRC. Experimental medicine [Internet]. 2019 [cited 2019 Jun 30]. Available from: https://mrc.ukri.org/research/initiatives/experimental-medicine/

75. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov. 2010;9(3):203–14.

76. Littman BH, Williams SA. The ultimate model organism: progress in experimental medicine. Nat Rev Drug Discov. 2005 Aug;4(8):631–8.

77. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001 Mar;69(3):89–95.

78. Abramson RG, Burton KR, Yu J-PJ, Scalzetti EM, Yankeelov TE, Rosenkrantz AB, et al. Methods and Challenges in Quantitative Imaging Biomarker Development. Acad Radiol. 2015 Jan;22(1):25–32.

79. Woodcock J, Woosley R. The FDA critical path initiative and its influence on new drug development. Annu Rev Med. 2008;59:1–12.

80. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource. Silver Spring (MD): Food and Drug Administration (US); 2016.

81. O'Connor JPB, Aboagye EO, Adams JE, Aerts HJWL, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. Nature Reviews Clinical Oncology. 2017 Mar;14(3):169–86.

82. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, et al. Metrology Standards for Quantitative Imaging Biomarkers. Radiology. 2015 Dec;277(3):813–25.

83. Shukla-Dave A, Obuchowski NA, Chenevert TL, Jambawalikar S, Schwartz LH, Malyarenko D, et al. Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. Journal of Magnetic Resonance Imaging. 2019;49(7):e101–21.

84. Brandt KD, Mazzuca SA. Lessons learned from nine clinical trials of disease-modifying osteoarthritis drugs. Arthritis Rheum. 2005 Nov;52(11):3349–59.

85. Karsdal MA, Michaelis M, Ladel C, Siebuhr AS, Bihlet AR, Andersen JR, et al. Disease-modifying treatments for osteoarthritis (DMOADs) of the knee and hip: lessons learned from failures and opportunities for the future. Osteoarthr Cartil. 2016 Dec;24(12):2013–21.

86. Bannuru RR, McAlindon TE, Sullivan MC, Wong JB, Kent DM, Schmid CH. Effectiveness and Implications of Alternative Placebo Treatments: A Systematic Review and Network Meta-analysis of Osteoarthritis Trials. Annals of Internal Medicine. 2015 Sep 1;163(5):365.

87. Kraus VB, Simon LS, Katz JN, Neogi T, Hunter D, Guermazi A, et al. Proposed study designs for approval based on a surrogate endpoint and a post-marketing confirmatory study under FDA's accelerated approval regulations for disease modifying osteoarthritis drugs. Osteoarthr Cartil. 2019 Apr;27(4):571–9.

88. Guermazi A, Alizai H, Crema MD, Trattnig S, Regatte RR, Roemer FW. Compositional MRI techniques for evaluation of cartilage degeneration in osteoarthritis. Osteoarthr Cartil. 2015 Oct;23(10):1639–53.

89. Mäkelä HI, Gröhn OH, Kettunen MI, Kauppinen RA. Proton exchange as a relaxation mechanism for T1 in the rotating frame in native and immobilized protein solutions. Biochem Biophys Res Commun. 2001 Dec 14;289(4):813–8.

90. Duvvuri U, Reddy R, Patel SD, Kaufman JH, Kneeland JB, Leigh JS. T1rho-relaxation in articular cartilage: effects of enzymatic degradation. Magn Reson Med. 1997 Dec;38(6):863–7.

91. Akella SVS, Reddy Regatte R, Gougoutas AJ, Borthakur A, Shapiro EM, Kneeland JB, et al. Proteoglycan-induced changes in T1ρ-relaxation of articular cartilage at 4T. Magnetic Resonance in Medicine. 2001;46(3):419–423.

92. Nieminen MT, Rieppo J, Töyräs J, Hakumäki JM, Silvennoinen J, Hyttinen MM, et al. T2 relaxation reveals spatial collagen architecture in articular cartilage: a comparative quantitative MRI and polarized light microscopic study. Magn Reson Med. 2001 Sep;46(3):487–93.

93. Nissi MJ, Rieppo J, Töyräs J, Laasanen MS, Kiviranta I, Jurvelin JS, et al. T2 relaxation time mapping reveals age- and species-related diversity of collagen network architecture in articular cartilage. Osteoarthr Cartil. 2006 Dec;14(12):1265–71.

94. Lüsse S, Claassen H, Gehrke T, Hassenpflug J, Schünke M, Heller M, et al. Evaluation of water content by spatially resolved transverse relaxation times of human articular cartilage. Magn Reson Imaging. 2000 May;18(4):423–30.

95. Bashir A, Gray ML, Boutin RD, Burstein D. Glycosaminoglycan in articular cartilage: in vivo assessment with delayed Gd(DTPA)(2-)-enhanced MR imaging. Radiology. 1997 Jan 11;205(2):551–8.

96. Burstein D, Velyvis J, Scott KT, Stock KW, Kim Y-J, Jaramillo D, et al. Protocol issues for delayed Gd(DTPA)2--enhanced MRI (dGEMRIC) for clinical evaluation of articular cartilage. Magn Reson Med. 2001 Jan 1;45(1):36–41.

97. van Tiel J, Kotek G, Reijman M, Bos PK, Bron EE, Klein S, et al. Is T1ρ Mapping an Alternative to Delayed Gadolinium-enhanced MR Imaging of Cartilage in the Assessment of Sulphated Glycosaminoglycan Content in Human Osteoarthritic Knees? An in Vivo Validation Study. Radiology. 2016 May;279(2):523–31.

98. Borthakur A, Shapiro EM, Beers J, Kudchodkar S, Kneeland JB, Reddy R. Sensitivity of MRI to proteoglycan depletion in cartilage: comparison of sodium and proton MRI. Osteoarthr Cartil. 2000 Jul 1;8(4):288–93.

99. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS med. 2009;6(7):e1000097.

100. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). Journal of Clinical Epidemiology. 2010 Aug;63(8):854–61.

101. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011;155(8):529–536.

102. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979 Mar;86(2):420–8.

103. Obuchowski NA, Barnhart HX, Buckler AJ, Pennello G, Wang X-F, Kalpathy-Cramer J, et al. Statistical issues in the comparison of quantitative imaging biomarker algorithms using pulmonary nodule volume as an example. Stat Methods Med Res. 2015 Feb;24(1):107–40.

104. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar;33(1):159–74.

105. The Nordic Cochrane Centre. Review Manager (RevMan). Copenhagen: The Cochrane Collaboration; 2014.

106. Anandacoomarasamy A, Leibman S, Smith G, Caterson I, Giuffre B, Fransen M, et al. Weight loss in obese people has structure-modifying effects on medial but not on lateral knee articular cartilage. Ann Rheum Dis. 2012 Jan 1;71(1):26–32.

107. Balamoody S, Williams TG, Wolstenholme C, Waterton JC, Bowes M, Hodgson R, et al. Magnetic resonance transverse relaxation time T2 of knee cartilage in osteoarthritis at 3-T: a cross-sectional multicentre, multivendor reproducibility study. Skeletal Radiol. 2013 Apr;42(4):511–20.

108. Baum T, Joseph GB, Arulanandan A, Nardo L, Virayavanich W, Carballido-Gamio J, et al. Association of magnetic resonance imaging–based knee cartilage T2 measurements and focal knee lesions with knee pain: Data from the Osteoarthritis Initiative. Arthritis Care Res. 2012 Feb 1;64(2):248–55.

109. Blumenkrantz G, Lindsey CT, Dunn TC, Jin H, Ries MD, Link TM, et al. A pilot, two-year longitudinal study of the interrelationship between trabecular bone and articular cartilage in the osteoarthritic knee. Osteoarthr Cartil. 2004 Dec;12(12):997–1005.

110. Bron EE, van Tiel J, Smit H, Poot DHJ, Niessen WJ, Krestin GP, et al. Image registration improves human knee cartilage T1 mapping with delayed gadolinium-enhanced MRI of cartilage (dGEMRIC). Eur Radiol. 2013 Jan;23(1):246–52.

111. Carballido-Gamio J., Link T.M., Majumdar S. New techniques for cartilage magnetic resonance imaging relaxation time analysis: Texture analysis of flattened cartilage and localized intra- and inter-subject comparisons. Magnetic Resonance in Medicine. 2008;59(6):1472–7.

112. Dardzinski BJ, Schneider E. Radiofrequency (RF) coil impacts the value and reproducibility of cartilage spin-spin (T2) relaxation time measurements. Osteoarthr Cartil. 2013 May;21(5):710–20.

113. Duryea J, Cheng C, Schaefer LF, Smith S, Madore B. Integration of accelerated MRI and post-processing software: a promising method for studies of knee osteoarthritis. Osteoarthr Cartil. 2016 Nov;24(11):1905–9.

114. Guha A, Wyatt C, Karampinos DC, Nardo L, Link TM, Majumdar S. Spatial variations in magnetic resonance-based diffusion of articular cartilage in knee osteoarthritis. Magn Reson Imaging. 2015 Nov;33(9):1051–8.

115. Gupta R, Virayavanich W, Kuo D, Su F, Link T, Ma B, et al. MR T(1)ρ quantification of cartilage focal lesions in acutely injured knees: correlation with arthroscopic evaluation. Magn Reson Imaging. 2014 Dec;32(10):1290–6.

116. Hada S, Kaneko H, Sadatsuki R, Liu L, Futami I, Kinoshita M, et al. The degeneration and destruction of femoral articular cartilage shows a greater degree of deterioration than that of the tibial and patellar articular cartilage in early stage knee osteoarthritis: a cross-sectional study. Osteoarthr Cartil. 2014 Oct;22(10):1583–9.

117. Hannila I, Räinä SS, Tervonen O, Ojala R, Nieminen MT. Topographical variation of T2 relaxation time in the young adult knee cartilage at 1.5 T. Osteoarthr Cartil. 2009 Dec;17(12):1570–5.

118. Hannila I, Lammentausta E, Tervonen O, Nieminen MT. The repeatability of T2 relaxation time measurement of human knee articular cartilage. Magn Reson Mater Phy. 2015 Jul 11;28(6):547–53.

119. Hesper T, Miese FR, Hosalkar HS, Behringer M, Zilkens C, Antoch G, et al. Quantitative T2(*) assessment of knee joint cartilage after running a marathon. Eur J Radiol. 2015 Feb;84(2):284–9.

120. Holtzman DJ, Theologis AA, Carballido-Gamio J, Majumdar S, Li X, Ma CB. T1ρ and T2 Quantitative Magnetic Resonance Imaging Analysis of Cartilage Regeneration Following Microfracture and Mosaicplasty Cartilage Resurfacing Procedures. J Magn Reson Imaging. 2010 Oct;32(4):914–23.

121. Hovis KK, Stehling C, Souza RB, Haughom BD, Baum T, Nevitt M, et al. Physical activity is associated with magnetic resonance imaging-based knee cartilage T2 measurements in asymptomatic subjects with and those without osteoarthritis risk factors. Arthritis Rheum. 2011 Aug;63(8):2248–56.

122. Hovis KK, Alizai H, Tham S-C, Souza RB, Nevitt MC, McCulloch CE, et al. Non-traumatic anterior cruciate ligament abnormalities and their relationship to osteoarthritis using morphological grading and cartilage T2 relaxation times: data from the Osteoarthritis Initiative (OAI). Skeletal Radiol. 2012 Nov;41(11):1435–43.

123. Jordan CD, McWalter EJ, Monu UD, Watkins RD, Chen W, Bangerter NK, et al. Variability of CubeQuant T1ρ, quantitative DESS T2, and cones sodium MRI in knee cartilage. Osteoarthr Cartil. 2014 Oct;22(10):1559–67.

124. Joseph GB, Baum T, Carballido-Gamio J, Nardo L, Virayavanich W, Alizai H, et al. Texture analysis of cartilage T 2 maps: individuals with risk factors for OA have higher and more heterogeneous knee cartilage MR T 2 compared to normal controls-data from the osteoarthritis initiative. Arthritis research & therapy. 2011;13(5):1.

125. Joseph GB, Baum T, Alizai H, Carballido-Gamio J, Nardo L, Virayavanich W, et al. Baseline mean and heterogeneity of MR cartilage T2 are associated with morphologic degeneration of cartilage, meniscus, and bone marrow over 3 years – data from the Osteoarthritis Initiative. Osteoarthr Cartil. 2012 Jul 1;20(7):727–35.

126. Juras V, Bohndorf K, Heule R, Kronnerwetter C, Szomolanyi P, Hager B, et al. A comparison of multi-echo spin-echo and triple-echo steady-state T2 mapping for in vivo evaluation of articular cartilage. Eur Radiol. 2016;26:1905–12.

127. Koli J, Multanen J, Kujala UM, Häkkinen A, Nieminen MT, Kautiainen H, et al. Effects of Exercise on Patellar Cartilage in Women with Mild Knee Osteoarthritis. Med Sci Sports Exerc. 2015 Sep;47(9):1767–74.

128. Li X, Wyatt C, Rivoire J, Han E, Chen W, Schooler J, et al. Simultaneous acquisition of T1ρ and T2 quantification in knee cartilage: Repeatability and diurnal variation. J Magn Reson Imaging. 2014 May 1;39(5):1287–93.

129. Li X, Han ET, Ma CB, Link TM, Newitt DC, Majumdar S. In vivo 3T spiral imaging based multi-slice T(1rho) mapping of knee cartilage in osteoarthritis. Magn Reson Med. 2005;54(4):929–36.

130. Li X, Pedoia V, Kumar D, Rivoire J, Wyatt C, Lansdown D, et al. Cartilage T1ρ and T2 relaxation times: longitudinal reproducibility and variations using different coils, MR systems and sites. Osteoarthr Cartil. 2015 Dec;23(12):2214–23.

131. Liebl H, Joseph G, Nevitt MC, Singh N, Heilmeier U, Subburaj K, et al. Early T2 changes predict onset of radiographic knee osteoarthritis: data from the osteoarthritis initiative. Ann Rheum Dis. 2015;74(7):1353–9.

132. Liess C, Lüsse S, Karger N, Heller M, Glüer C-C. Detection of changes in cartilage water content using MRI T2-mapping in vivo. Osteoarthr Cartil. 2002 Dec;10(12):907–13.

133. Liu F, Choi KW, Samsonov A, Spencer RG, Wilson JJ, Block WF, et al. Articular Cartilage of the Human Knee Joint: In Vivo Multicomponent T2 Analysis at 3.0 T. Radiology. 2015 Nov;277(2):477–88.

134. Madelin G, Babb JS, Xia D, Chang G, Jerschow A, Regatte RR. Reproducibility and repeatability of quantitative sodium magnetic resonance imaging in vivo in articular cartilage at 3 T and 7 T. Magn Reson Med. 2012 Sep 1;68(3):841–9.

135. Matsubara H, Okazaki K, Takayama Y, Osaki K, Matsuo Y, Honda H, et al. Detection of early cartilage deterioration associated with meniscal tear using T1ρ mapping magnetic resonance imaging. BMC Musculoskelet Disord. 2015;16:22.

136. Mosher TJ, Zhang Z, Reddy R, Boudhar S, Milestone BN, Morrison WB, et al. Knee Articular Cartilage Damage in Osteoarthritis: Analysis of MR Image Biomarker Reproducibility in ACRIN-PA 4001 Multicenter Trial. Radiology. 2011 Mar 1;258(3):832–42.

137. Mosher TJ, Liu Y, Yang QX, Yao J, Smith R, Dardzinski BJ, et al. Age dependency of cartilage magnetic resonance imaging T2 relaxation times in asymptomatic women. Arthritis Rheum. 2004 Sep 1;50(9):2820–8.

138. Multanen J, Rauvala E, Lammentausta E, Ojala R, Kiviranta I, Häkkinen A, et al. Reproducibility of imaging human knee cartilage by delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) at 1.5 Tesla. Osteoarthr Cartil. 2009 May;17(5):559–64.

139. Multanen J, Heinonen A, Häkkinen A, Kautiainen H, Kujala UM, Lammentausta E, et al. Bone and cartilage characteristics in postmenopausal women with mild knee radiographic osteoarthritis and those without radiographic osteoarthritis. J Musculoskelet Neuronal Interact. 2015 Mar;15(1):69–77.

140. Newbould RD, Miller SR, Toms LD, Swann P, Tielbeek JAW, Gold GE, et al. T2* measurement of the knee articular cartilage in osteoarthritis at 3T. J Magn Reson Imaging. 2012;35(6):1422–9.

141. Newbould RD, Miller SR, Tielbeek JAW, Toms LD, Rao AW, Gold GE, et al. Reproducibility of sodium MRI measures of articular cartilage of the knee in osteoarthritis. Osteoarthr Cartil. 2012 Jan;20(1):29–35.

142. Nishioka H, Hirose J, Nakamura E, Oniki Y, Takada K, Yamashita Y, et al. T1ρ and T2 mapping reveal the in vivo extracellular matrix of articular cartilage. J Magn Reson Imaging. 2012 Jan;35(1):147–55.

143. Nishioka H, Hirose J, Nakamura E, Okamoto N, Karasugi T, Taniwaki T, et al. Detecting ICRS grade 1 cartilage lesions in anterior cruciate ligament injury using T1ρ and T2 mapping. European Journal of Radiology. 2013 Sep 1;82(9):1499–505.

144. Nishioka H, Hirose J, Okamoto N, Okada T, Oka K, Taniwaki T, et al. Evaluation of the relationship between T1ρ and T2 values and patella cartilage degeneration in patients of the same age group. Eur J Radiol. 2015 Mar;84(3):463–8.

145. Pan J, Pialat J-B, Joseph T, Kuo D, Joseph GB, Nevitt MC, et al. Knee cartilage T2 characteristics and evolution in relation to morphologic abnormalities detected at 3-T MR imaging: a longitudinal study of the normal control cohort from the Osteoarthritis Initiative. Radiology. 2011 Nov;261(2):507–15.

146. Pedoia V, Li X, Su F, Calixto N, Majumdar S. Fully automatic analysis of the knee articular cartilage T1ρ relaxation time using voxel-based relaxometry. J Magn Reson Imaging. 2016 Apr;43(4):970–80.

147. Raya JG, Dettmann E, Notohamiprodjo M, Krasnokutsky S, Abramson S, Glaser C. Feasibility of in vivo diffusion tensor imaging of articular cartilage with coverage of all cartilage regions. Eur Radiol. 2014 Jul;24(7):1700–6.

148. Raya JG, Horng A, Dietrich O, Krasnokutsky S, Beltran LS, Storey P, et al. Articular Cartilage: In Vivo Diffusion-Tensor Imaging. Radiology. 2012 Feb 1;262(2):550–9.

149. Schleich C, Bittersohl B, Miese F, Schmitt B, Müller-Lutz A, Sondern M, et al. Glycosaminoglycan chemical exchange saturation transfer at 3T MRI in asymptomatic knee joints. Acta Radiol. 2016 May;57(5):627–32.

150. Serebrakian AT, Poulos T, Liebl H, Joseph GB, Lai A, Nevitt MC, et al. Weight loss over 48 months is associated with reduced progression of cartilage T2 relaxation time values: Data from the osteoarthritis initiative. J Magn Reson Imaging. 2015 May 1;41(5):1272–80.

151. Singh A, Haris M, Cai K, Kogan F, Hariharan H, Reddy R. High resolution T1ρ mapping of in vivo human knee cartilage at 7T. PLoS ONE. 2014;9(5):e97486.

152. Siversson C, Tiderius C-J, Neuman P, Dahlberg L, Svensson J. Repeatability of T1-quantification in dGEMRIC for three different acquisition techniques: two-dimensional inversion recovery, three-dimensional look locker, and three-dimensional variable flip angle. J Magn Reson Imaging. 2010 May;31(5):1203–9.

153. Sritanyaratana N, Samsonov A, Mossahebi P, Wilson JJ, Block WF, Kijowski R. Cross-relaxation imaging of human patellar cartilage in vivo at 3.0T. Osteoarthr Cartil. 2014 Oct;22(10):1568–76.

154. Stehling C, Baum T, Mueller-Hoecker C, Liebl H, Carballido-Gamio J, Joseph GB, et al. A novel fast knee cartilage segmentation technique for T2 measurements at MR imaging--data from the Osteoarthritis Initiative. Osteoarthr Cartil. 2011 Aug;19(8):984–9.

155. Stehling C, Liebl H, Krug R, Lane NE, Nevitt MC, Lynch J, et al. Patellar Cartilage: T2 Values and Morphologic Abnormalities at 3.0-T MR Imaging in Relation to Physical Activity in Asymptomatic Subjects from the Osteoarthritis Initiative. Radiology. 2010 Jan 7;254(2):509–20.

156. Surowiec RK, Lucas EP, Fitzcharles EK, Petre BM, Dornan GJ, Giphart JE, et al. T2 values of articular cartilage in clinically relevant subregions of the asymptomatic knee. Knee Surg Sports Traumatol Arthrosc. 2014 Jun;22(6):1404–14.

157. Takayama Y, Hatakenaka M, Tsushima H, Okazaki K, Yoshiura T, Yonezawa M, et al. T1ρ is superior to T2 mapping for the evaluation of articular cartilage denaturalization with osteoarthritis: radiological-pathological correlation after total knee arthroplasty. Eur J Radiol. 2013;82(4):e192-8.

158. Tiderius CJ, Tjörnstrand J, Åkeson P, Södersten K, Dahlberg L, Leander P. Delayed gadolinium-enhanced MRI of cartilage (dGEMRIC): intra- and interobserver variability in standardized drawing of regions of interest. Acta Radiol. 2004 Jan 10;45(6):628–34.

159. van Tiel J, Bron EE, Tiderius CJ, Bos PK, Reijman M, Klein S, et al. Reproducibility of 3D delayed gadolinium enhanced MRI of cartilage (dGEMRIC) of the knee at 3.0 T in patients with early stage osteoarthritis. Eur Radiol. 2013 Feb;23(2):496–504.

160. Welsch GH, Apprich S, Zbyn S, Mamisch TC, Mlynarik V, Scheffler K, et al. Biochemical (T2, T2* and magnetisation transfer ratio) MRI of knee cartilage: feasibility at ultra-high field (7T) compared with high field (3T) strength. Eur Radiol. 2010 Dec 12;21(6):1136–43.

161. Wiener E, Pfirrmann CWA, Hodler J. Spatial variation in T1 of healthy human articular cartilage of the knee joint. Br J Radiol. 2010 Jan 6;83(990):476–85.

162. Williams A, Qian Y, Chu CR. UTE-T2* mapping of human articular cartilage in vivo: a repeatability assessment. Osteoarthr Cartil. 2011 Jan;19(1):84–8.

163. Zuo J, Li X, Banerjee S, Han E, Majumdar S. Parallel imaging of knee cartilage at 3 Tesla. J Magn Reson Imaging. 2007;26(4):1001–9.

164. Dunn TC, Lu Y, Jin H, Ries MD, Majumdar S. T2 Relaxation Time of Cartilage at MR Imaging: Comparison with Severity of Knee Osteoarthritis. Radiology. 2004 Jan 8;232(2):592–8.

165. Eckstein F, Le Graverand MPH, Charles HC, Hunter DJ, Kraus VB, Sunyer T, et al. Clinical, radiographic, molecular and MRI-based predictors of cartilage loss in knee osteoarthritis. Ann Rheum Dis. 2011 Jul;70(7):1223–30.

166. Li W, Scheidegger R, Wu Y, Edelman RR, Farley M, Krishnan N, et al. Delayed contrast-enhanced MRI of cartilage: Comparison of nonionic and ionic contrast agents. Magn Reson Med. 2010 Nov 1;64(5):1267–73.

167. Li X, Pai A, Blumenkrantz G, Carballido-Gamio J, Link T, Ma B, et al. Spatial distribution and relationship of T1rho and T2 relaxation times in knee cartilage with osteoarthritis. Magn Reson Med. 2009;61(6):1310–8.

168. Li X, Ma CB, Link TM, Castillo D-D, Blumenkrantz G, Lozano J, et al. In vivo T1rho and T2 mapping of articular cartilage in osteoarthritis of the knee using 3 Tesla MRI. Osteoarthr Cartil. 2007 Jul;15(7):789–97.

169. Madelin G, Babb J, Xia D, Chang G, Krasnokutsky S, Abramson SB, et al. Articular cartilage: evaluation with fluid-suppressed 7.0-T sodium MR imaging in subjects with and subjects without osteoarthritis. Radiology. 2013 Aug;268(2):481–91.

170. Owman H, Ericsson YB, Englund M, Tiderius CJ, Tjörnstrand J, Roos EM, et al. Association between delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) and joint space narrowing and osteophytes: a cohort study in patients with partial meniscectomy with 11 years of follow-up. Osteoarthr Cartil. 2014 Oct 1;22(10):1537–41.

171. Souza RB, Kumar D, Calixto N, Singh J, Schooler J, Subburaj K, et al. Response of knee cartilage T1rho and T2 relaxation times to in vivo mechanical loading in individuals with and without knee osteoarthritis. Osteoarthr Cartil. 2014 Oct;22(10):1367–76.

172. Souza RB, Feeley BT, Zarins ZA, Link TM, Li X, Majumdar S. T1rho MRI relaxation in knee OA subjects with varying sizes of cartilage lesions. Knee. 2013 Mar 1;20(2):113–9.

173. Stahl R, Blumenkrantz G, Carballido-Gamio J, Zhao S, Munoz T, Hellio Le Graverand-Gastineau MP, et al. MRI-derived T2 relaxation times and cartilage morphometry of the tibio-femoral joint in subjects with and without osteoarthritis during a 1-year follow-up. Osteoarthr Cartil. 2007 Nov;15(11):1225–34.

174. Stahl R, Luke A, Li X, Carballido-Gamio J, Ma CB, Majumdar S, et al. T1rho, T2 and focal knee cartilage abnormalities in physically active and sedentary healthy subjects versus early OA patients--a 3.0-Tesla MRI study. Eur Radiol. 2009;19(1):132–43.

175. Wang L, Regatte RR. Quantitative mapping of human cartilage at 3.0T: parallel changes in T2, T1p, and dGEMRIC. Acad Radiol. 2014;21(4):463–71.

176. Wirth W, Maschek S, Roemer FW, Eckstein F. Layer-specific femorotibial cartilage T2 relaxation time in knees with and without early knee osteoarthritis: Data from the Osteoarthritis Initiative (OAI). Sci Rep. 2016 Sep 27;6:34202.

177. Wyatt C, Guha A, Venkatachari A, Li X, Krug R, Kelley DE, et al. Improved differentiation between knees with cartilage lesions and controls using 7T

relaxation time mapping. Journal of Orthopaedic Translation. 2015 Oct;3(4):197–204.

178. Newbould RD, Miller SR, Upadhyay N, Rao AW, Swann P, Gold GE, et al. T1-weighted sodium MRI of the articulator cartilage in osteoarthritis: a cross sectional and longitudinal study. PLoS ONE. 2013;8(8):e73067.

179. Yao W, Qu N, Lu Z, Yang S. The application of T1 and T2 relaxation time and magnetization transfer ratios to the early diagnosis of patellar cartilage osteoarthritis. Skeletal Radiol. 2009 Nov;38(11):1055–62.

180. Athelogou M, Kim HJ, Dima A, Obuchowski N, Peskin A, Gavrielides MA, et al. Algorithm Variability in the Estimation of Lung Nodule Volume From Phantom CT Scans: Results of the QIBA 3A Public Challenge. Acad Radiol. 2016 Aug;23(8):940–52.

181. Ng CS, Raunig DL, Jackson EF, Ashton EA, Kelcz F, Kim KB, et al. Reproducibility of perfusion parameters in dynamic contrast-enhanced MRI of lung and liver tumors: effect on estimates of patient sample size in clinical trials and on individual patient responses. AJR Am J Roentgenol. 2010 Feb;194(2):W134-140.

182. Regatte RR, Akella SVS, Lonner J h., Kneeland J b., Reddy R. T1ρ relaxation mapping in human osteoarthritis (OA) cartilage: Comparison of T1ρ with T2. Journal of Magnetic Resonance Imaging. 2006;23(4):547–553.

183. Smith HE, Mosher TJ, Dardzinski BJ, Collins BG, Collins CM, Yang QX, et al. Spatial variation in cartilage T2 of the knee. Journal of Magnetic Resonance Imaging. 2001 Jul;14(1):50–5.

184. Jungmann PM, Kraus MS, Nardo L, Liebl H, Alizai H, Joseph GB, et al. T(2) relaxation time measurements are limited in monitoring progression, once advanced cartilage defects at the knee occur: longitudinal data from the osteoarthritis initiative. J Magn Reson Imaging. 2013;38(6):1415–24.

185. Turmezei T, Poole K, Treece G. Three-dimensional measurement of the hip joint space from clinical CT for the assessment and prediction of osteoarthritis. The Lancet. 2016;387:S101.

186. Nebelung S, Sondern B, Oehrl S, Tingart M, Rath B, Pufe T, et al. Functional MR Imaging Mapping of Human Articular Cartilage Response to Loading. Radiology. 2016 Aug 26;282(2):464–74.

187. Monu UD, Jordan CD, Samuelson BL, Hargreaves BA, Gold GE, McWalter EJ. Cluster analysis of quantitative MRI T2 and T1ρ relaxation times of cartilage identifies differences between healthy and ACL-injured individuals at 3T. Osteoarthr Cartil. 2017 Apr;25(4):513–20.

188. Jørgensen DR, Lillholm M, Genant HK, Dam EB. On Subregional Analysis of Cartilage Loss from Knee MRI. Cartilage. 2013 Apr;4(2):121–30.

189. Williams TG, Holmes AP, Waterton JC, Maciewicz RA, Hutchinson CE, Moots RJ, et al. Anatomically corresponded regional analysis of cartilage in asymptomatic and osteoarthritic knees by statistical shape modelling of the bone. IEEE Trans Med Imaging. 2010 Aug;29(8):1541–59.

190. Favre J, Erhart-Hledik JC, Blazek K, Fasel B, Gold GE, Andriacchi TP. Anatomically Standardized Maps Reveal Distinct Patterns of Cartilage Thickness With Increasing Severity of Medial Compartment Knee Osteoarthritis. J Orthop Res. 2017 Nov;35(11):2442–51.

191. Poole KES, Treece GM, Mayhew PM, Vaculík J, Dungl P, Horák M, et al. Cortical Thickness Mapping to Identify Focal Osteoporosis in Patients with Hip Fracture. PLoS One. 2012 Jun 11;7(6):e38466.

192. Treece G, Gee A. Cortical Bone Mapping: Measurement and Statistical Analysis of Localised Skeletal Changes. Curr Osteoporos Rep. 2018 Oct;16(5):617–25.

193. Treece GM, Gee AH, Mayhew PM, Poole KES. High resolution cortical bone thickness measurement from clinical CT data. Med Image Anal. 2010 Jun;14(3):276–90.

194. Pearson RA, Treece GM. Measurement of the bone endocortical region using clinical CT. Med Image Anal. 2018;44:28–40.

195. Treece GM, Gee AH. Independent measurement of femoral cortical thickness and cortical bone density using clinical CT. Med Image Anal. 2015 Feb;20(1):249–64.

196. Whitmarsh T, Treece GM, Gee AH, Poole KE. Mapping Bone Changes at the Proximal Femoral Cortex of Postmenopausal Women in Response to Alendronate and Teriparatide Alone, Combined or Sequentially. Journal of Bone and Mineral Research. 2015 Jul 1;30(7):1309–18.

197. Turmezei TD, Treece GM, Gee AH, Houlden R, Poole KES. A new quantitative 3D approach to imaging of structural joint disease. Scientific Reports. 2018 Jun 18;8(1):9280.

198. MacKay JW, Turmezei TD, Kaggie J, Morgan AR, Janiczek RL, Khan W, et al. Repeatability and discrimination validity of cartilage imaging biomarkers for experimental medicine studies of knee osteoarthritis. Osteoarthr Cartil. 2018 Apr;26:S464–5.

199. Treece GM, Prager RW, Gee AH. Regularised Marching Tetrahedra: Improved Iso-Surface Extraction. Computers and Graphics. 1998;23:583–598.

200. Netter F. Atlas of Human Anatomy. 6th edition. Philadelphia, PA: Saunders; 2014. p. 640

201. Norman B, Pedoia V, Majumdar S. Use of 2D U-Net Convolutional Neural Networks for Automated Cartilage and Meniscus Segmentation of Knee MR Imaging Data to Determine Relaxometry and Morphometry. Radiology. 2018 Mar 27;172322.

202. Dam EB, Lillholm M, Marques J, Nielsen M. Automatic segmentation of high- and low-field knee MRIs using knee image quantification with data from the osteoarthritis initiative. J Med Imaging. 2015 Apr;2(2):024001.

203. Guillard G, Vincent G, Brett A, Conaghan P, Bowes M. Automated Segmentation of Cartilage Provides Comparable Accuracy and Better Responsiveness Than Manual Segmentation: Data from the Osteoarthritis Initiative. Arthritis Rheumatol. 2016;68(S10).

204. Goldring MB, Goldring SR. Articular cartilage and subchondral bone in the pathogenesis of osteoarthritis. Ann N Y Acad Sci. 2010 Mar;1192:230–7.

205. Kwan Tat S, Lajeunesse D, Pelletier J-P, Martel-Pelletier J. Targeting subchondral bone for treating osteoarthritis: what is the evidence? Best Practice & Research Clinical Rheumatology. 2010 Feb;24(1):51–70.

206. Chang G, Xia D, Chen C, Madelin G, Abramson SB, Babb JS, et al. 7T MRI detects deterioration in subchondral bone microarchitecture in subjects with mild knee osteoarthritis as compared with healthy controls. J Magn Reson Imaging. 2015 May;41(5):1311–7.

207. Kraus VB, Feng S, Wang S, White S, Ainslie M, Graverand M-PHL, et al. Subchondral Bone Trabecular Integrity Predicts and Changes Concurrently With Radiographic and Magnetic Resonance Imaging-Determined Knee Osteoarthritis Progression. Arthritis Rheum. 2013 Jul;65(7):1812–21.

208. Lo GH, Schneider E, Driban JB, Price LL, Hunter DJ, Eaton CB, et al. Periarticular bone predicts knee osteoarthritis progression: Data from the Osteoarthritis Initiative. Semin Arthritis Rheum. 2018;48(2):155–61.

209. Lo GH, Tassinari AM, Driban JB, Price LL, Schneider E, Majumdar S, et al. Cross-sectional DXA and MR measures of tibial periarticular bone associate with radiographic knee osteoarthritis severity. Osteoarthr Cartil. 2012 Jul;20(7):686–93.

210. Hirvasniemi J, Thevenot J, Guermazi A, Podlipská J, Roemer FW, Nieminen MT, et al. Differences in tibial subchondral bone structure evaluated using plain

radiographs between knees with and without cartilage damage or bone marrow lesions - the Oulu Knee Osteoarthritis study. Eur Radiol. 2017 Nov;27(11):4874–82.

211. MacKay JW, Murray PJ, Kasmai B, Johnson G, Donell ST, Toms AP. MRI texture analysis of subchondral bone at the tibial plateau. Eur Radiol. 2016 Sep;26(9):3034–45.

212. MacKay JW, Murray PJ, Low SBL, Kasmai B, Johnson G, Donell ST, et al. Quantitative analysis of tibial subchondral bone: Texture analysis outperforms conventional trabecular microarchitecture analysis. J Magn Reson Imaging. 2016 May;43(5):1159–70.

213. MacKay JW, Murray PJ, Kasmai B, Johnson G, Donell ST, Toms AP. Subchondral bone in osteoarthritis: association between MRI texture analysis and histomorphometry. Osteoarthr Cartil. 2017 May;25(5):700–7.

214. Hunter DJ, Nevitt M, Losina E, Kraus V. Biomarkers for osteoarthritis: Current position and steps towards further validation. Best Practice & Research Clinical Rheumatology. 2014 Feb;28(1):61–71.

215. Nevitt MC, Peterfy C, Guermazi A, Felson DT, Duryea J, Woodworth T, et al. Longitudinal performance evaluation and validation of fixed-flexion radiography of the knee for detection of joint space loss. Arthritis Rheum. 2007 May;56(5):1512–20.

216. Duryea J, Neumann G, Niu J, Totterman S, Tamez J, Dabrowski C, et al. Comparison of radiographic joint space width with magnetic resonance imaging cartilage morphometry: analysis of longitudinal data from the Osteoarthritis Initiative. Arthritis Care Res . 2010 Jul;62(7):932–7.

217. Felson DT, Nevitt MC, Yang M, Clancy M, Niu J, Torner JC, et al. A new approach yields high rates of radiographic progression in knee osteoarthritis. J Rheumatol. 2008 Oct;35(10):2047–54.

218. Iranpour-Boroujeni T, Li J, Lynch JA, Nevitt M, Duryea J, OAI Investigators. A new method to measure anatomic knee alignment for large studies of OA: data from the osteoarthritis initiative. Osteoarthr Cartil. 2014 Oct;22(10):1668–74.

219. Schneider E, Lo GH, Sloane G, Fanella L, Hunter DJ, Eaton CB, et al. Magnetic resonance imaging evaluation of weight-bearing subchondral trabecular bone in the knee. Skeletal Radiol. 2011 Jan 1;40(1):95–103.

220. Szczypiński PM, Strzelecki M, Materka A, Klepaczko A. MaZda--a software package for image texture analysis. Comput Methods Programs Biomed. 2009 Apr;94(1):66–76.

221. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. Stat Med. 2012 Dec 20;31(29):3972–81.

222. Ho D, Imai K, King G, Stuart EA. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. Journal of Statistical Software. 2011 Jun 14;42(8):1–28.

223. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software. 2010;33(1):1–22.

224. MacKay JW, Low SBL, Murray PJ, Kasmai B, Johnson G, Donell ST, et al. MR texture analysis of subchondral bone in osteoarthritis. In: Proc Intl Soc Mag Reson Med. Singapore; 2016. p. 4485.

225. LaValley MP, Lo GH, Price LL, Driban JB, Eaton CB, McAlindon TE. Development of a clinical prediction algorithm for knee osteoarthritis structural progression in a cohort study: value of adding measurement of subchondral bone density. Arthritis Res Ther. 2017 May 16;19(1):95.

226. Hunter D, Nevitt M, Lynch J, Kraus VB, Katz JN, Collins JE, et al. Longitudinal validation of periarticular bone area and 3D shape as biomarkers for knee OA progression? Data from the FNIH OA Biomarkers Consortium. Ann Rheum Dis. 2016 Sep;75(9):1607–14.

227. Hodgdon T, McInnes MDF, Schieda N, Flood TA, Lamb L, Thornhill RE. Can Quantitative CT Texture Analysis be Used to Differentiate Fat-poor Renal Angiomyolipoma from Renal Cell Carcinoma on Unenhanced CT Images? Radiology. 2015 Apr 23;276(3):787–96.

228. Ng F, Ganeshan B, Kozarski R, Miles KA, Goh V. Assessment of Primary Colorectal Cancer Heterogeneity by Using Whole-Tumor Texture Analysis: Contrast-enhanced CT Texture as a Biomarker of 5-year Survival. Radiology. 2013 Jan 1;266(1):177–84.

229. Rajani NK, Joshi NV, Elkhawad M, Melville A, Chowdhury M, Ganeshan B, et al. CT textural analysis of abdominal aortic aneurysms as a biomarker for aneurysm growth. The Lancet. 2014 Feb;383:S87.

230. O'Neill TW, Parkes MJ, Maricar N, Marjanovic EJ, Hodgson R, Gait AD, et al. Synovial tissue volume: a treatment target in knee osteoarthritis (OA). Ann Rheum Dis. 2016 Jan 1;75(1):84–90.

231. Eckstein F, Ateshian G, Burgkart R, Burstein D, Cicuttini F, Dardzinski B, et al. Proposal for a nomenclature for magnetic resonance imaging based measures of articular cartilage in osteoarthritis. Osteoarthr Cartil. 2006 Oct;14(10):974–83.

232. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum. 1986 Aug;29(8):1039–49.

233. Kothari M, Guermazi A, von Ingersleben G, Miaux Y, Sieffert M, Block JE, et al. Fixed-flexion radiography of the knee provides reproducible joint space width measurements in osteoarthritis. European Radiology. 2004 Sep;14(9):1568–73.

234. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther. 1998 Aug;28(2):88–96.

235. Waterton JC, Solloway S, Foster JE, Keen MC, Gandy S, Middleton BJ, et al. Diurnal variation in the femoral articular cartilage of the knee in young adult humans. Magn Reson Med. 2000 Jan;43(1):126–32.

236. Eckstein F, Tieschky M, Faber SC, Haubner M, Kolem H, Englmeier KH, et al. Effect of physical exercise on cartilage volume and thickness in vivo: MR imaging study. Radiology. 1998 Apr;207(1):243–8.

237. Subburaj K, Kumar D, Souza RB, Alizai H, Li X, Link TM, et al. The acute effect of running on knee articular cartilage and meniscus magnetic resonance relaxation times in young healthy adults. Am J Sports Med. 2012 Sep;40(9):2134–41.

238. Roberts C, Hughes S, Naish JH, Holliday K, Watson Y, Cheung S, et al. Use of An Individually Measured Hematocrit in DCE-MRI studies. In: Proc Intl Soc Mag Reson Med. Montreal; 2011. p. 1078.

239. Ross PD, Huang C, Karpf D, Lydick E, Coel M, Hirsch L, et al. Blinded reading of radiographs increases the frequency of errors in vertebral fracture detection. J Bone Miner Res. 1996 Nov;11(11):1793–800.

240. Bruynesteyn K, Van Der Heijde D, Boers M, Saudan A, Peloso P, Paulus H, et al. Detecting radiological changes in rheumatoid arthritis that are considered important by clinical experts: influence of reading with or without known sequence. J Rheumatol. 2002 Nov;29(11):2306–12.

241. Felson DT, Nevitt MC. Blinding images to sequence in osteoarthritis: evidence from other diseases. Osteoarthr Cartil. 2009 Mar 1;17(3):281–3.

242. Collins NJ, Prinsen C a. C, Christensen R, Bartels EM, Terwee CB, Roos EM. Knee Injury and Osteoarthritis Outcome Score (KOOS): systematic review and meta-analysis of measurement properties. Osteoarthr Cartil. 2016 Aug;24(8):1317–29.

243. Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, et al. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthr Cartil. 2004 May;12(5):389–99.

244. Runhaar J, Schiphof D, van Meer B, Reijman M, Bierma-Zeinstra SMA, Oei EHG. How to define subregional osteoarthritis progression using semi-quantitative MRI osteoarthritis knee score (MOAKS). Osteoarthr Cartil. 2014 Oct;22(10):1533–6.

245. Paradowski PT, Bergman S, Sundén-Lundius A, Lohmander LS, Roos EM. Knee complaints vary with age and gender in the adult population. Population-based reference data for the Knee injury and Osteoarthritis Outcome Score (KOOS). BMC Musculoskelet Disord. 2006 May 2;7:38.

246. Benichou OD, Hunter DJ, Nelson DR, Guermazi A, Eckstein F, Kwoh K, et al. One-year change in radiographic joint space width in patients with unilateral joint space narrowing: data from the Osteoarthritis Initiative. Arthritis Care Res . 2010 Jul;62(7):924–31.

247. Magnusson K, Kumm J, Turkiewicz A, Englund M. A naturally aging knee, or development of early knee osteoarthritis? Osteoarthr Cartil. 2018 Nov 1;26(11):1447–52.

248. Patterson BE, Culvenor AG, Barton CJ, Guermazi A, Stefanik JJ, Morris HG, et al. Worsening Knee Osteoarthritis Features on Magnetic Resonance Imaging 1 to 5 Years After Anterior Cruciate Ligament Reconstruction. Am J Sports Med. 2018 Oct;46(12):2873–83.

249. van Meer BL, Oei EHG, Meuffels DE, van Arkel ERA, Verhaar JAN, Bierma-Zeinstra SMA, et al. Degenerative Changes in the Knee 2 Years After Anterior Cruciate Ligament Rupture and Related Risk Factors: A Prospective Observational Follow-up Study. Am J Sports Med. 2016 Jun;44(6):1524–33.

250. Hill CL, Hunter DJ, Niu J, Clancy M, Guermazi A, Genant H, et al. Synovitis detected on magnetic resonance imaging and its relation to pain and cartilage loss in knee osteoarthritis. Annals of the Rheumatic Diseases. 2007 Dec 1;66(12):1599–603.

251. Zhang Y, Nevitt M, Niu J, Lewis C, Torner J, Guermazi A, et al. Fluctuation of knee pain and changes in bone marrow lesions, effusions, and synovitis on magnetic resonance imaging. Arthritis Rheum. 2011 Mar;63(3):691–9.

252. Miller TT, Staron RB, Feldman F, Cepel E. Meniscal position on routine MR imaging of the knee. Skeletal Radiol. 1997 Jul;26(7):424–7.

253. Patel R, Eltgroth M, Souza RB, Zhang CA, Majumdar S, Link TM, et al. Loaded versus unloaded magnetic resonance imaging (MRI) of the knee: Effect on meniscus extrusion in healthy volunteers and patients with osteoarthritis. European Journal of Radiology Open. 2016 Jan 1;3:100–7.

254. Puig L, Monllau JC, Corrales M, Pelfort X, Melendo E, Cáceres E. Factors affecting meniscal extrusion: correlation with MRI, clinical, and arthroscopic findings. Knee Surg Sports Traumatol Arthrosc. 2006 Apr;14(4):394–8.

255. Roemer FW, Kassim Javaid M, Guermazi A, Thomas M, Kiran A, Keen R, et al. Anatomical distribution of synovitis in knee osteoarthritis and its association with joint effusion assessed on non-enhanced and contrast-enhanced MRI. Osteoarthr Cartil 2010 Oct;18(10):1269–74.

256. Loeuille D., Chary-Valckenaere I., Champigneulle J., Rat A.-C., Toussaint F., Pinzano-Watrin A., et al. Macroscopic and microscopic features of synovial membrane inflammation in the osteoarthritic knee: Correlating magnetic resonance imaging findings with disease severity. Arthritis and Rheumatism. 2005.

257. Baker K, Grainger A, Niu J, Clancy M, Guermazi A, Crema M, et al. Relation of synovitis to knee pain using contrast-enhanced MRIs. Annals of the Rheumatic Diseases. 2010 Oct 1;69(10):1779–83.

258. Roemer FW, Kwoh CK, Hannon MJ, Hunter DJ, Eckstein F, Fujii T, et al. What comes first? Multitissue involvement leading to radiographic osteoarthritis: magnetic resonance imaging-based trajectory analysis over four years in the osteoarthritis initiative. Arthritis Rheum. 2015 May;67(8):2085–96.

259. Atukorala I, Kwoh CK, Guermazi A, Roemer FW, Boudreau RM, Hannon MJ, et al. Synovitis in knee osteoarthritis: a precursor of disease? Ann Rheum Dis. 2016 Feb;75(2):390–5.

260. Smith MD, Triantafillou S, Parker A, Youssef PP, Coleman M. Synovial membrane inflammation and cytokine production in patients with early osteoarthritis. J Rheumatol. 1997 Feb;24(2):365–71.

261. Waterton JC, Ho M, Nordenmark LH, Jenkins M, DiCarlo J, Guillard G, et al. Repeatability and response to therapy of dynamic contrast-enhanced magnetic resonance imaging biomarkers in rheumatoid arthritis in a large multicentre trial setting. European Radiology. 2017 Sep;27(9):3662–8.

262. Conaghan PG, Østergaard M, Bowes MA, Wu C, Fuerst T, van der Heijde D, et al. Comparing the effects of tofacitinib, methotrexate and the combination, on bone marrow oedema, synovitis and bone erosion in methotrexate-naive, early active

rheumatoid arthritis: results of an exploratory randomised MRI study incorporating semiquantitative and quantitative techniques. Ann Rheum Dis. 2016 Jun;75(6):1024–33.

263. Xanthopoulos E, Hutchinson CE, Adams JE, Bruce IN, Nash AFP, Holmes AP, et al. Improved wrist pannus volume measurement from contrast-enhanced MRI in rheumatoid arthritis using shuffle transform. Magn Reson Imaging. 2007 Jan;25(1):110–6.

264. Otsu N. A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics. 1979 Jan;9(1):62–6.

265. Perry TA, Gait A, O'Neill TW, Parkes MJ, Hodgson R, Callaghan MJ, et al. Measurement of synovial tissue volume in knee osteoarthritis using a semiautomated MRI-based quantitative approach. Magn Reson Med. 2019 May;81(5):3056–64.

266. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage. 2011 Feb 1;54(3):2033–44.

267. Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, et al. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. J Magn Reson Imaging. 1999 Sep;10(3):223–32.

268. Parker GJM, Roberts C, Macdonald A, Buonaccorsi GA, Cheung S, Buckley DL, et al. Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI. Magn Reson Med. 2006 Nov;56(5):993–1000.

269. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, et al. Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment. Statistical Methods in Medical Research. 2015 Feb;24(1):27–67.

270. Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow up measurements. BMJ. 2001 Nov 10;323(7321):1123–4.

271. Roberts C, Issa B, Stone A, Jackson A, Waterton JC, Parker GJM. Comparative study into the robustness of compartmental modeling and model-free analysis in DCE-MRI studies. J Magn Reson Imaging. 2006 Apr;23(4):554–63.

272. Galbraith SM, Lodge MA, Taylor NJ, Rustin GJS, Bentzen S, Stirling JJ, et al. Reproducibility of dynamic contrast-enhanced MRI in human muscle and

tumours: comparison of quantitative and semi-quantitative analysis. NMR Biomed. 2002 Apr;15(2):132–42.

273. Parry EL, Thomas MJ, Peat G. Defining acute flares in knee osteoarthritis: a systematic review. BMJ Open. 2018;8(7)

274. Guermazi A, Roemer FW, Hayashi D, Crema MD, Niu J, Zhang Y, et al. Assessment of synovitis with contrast-enhanced MRI using a whole-joint semiquantitative scoring system in people with, or at high risk of, knee osteoarthritis: the MOST study. Ann Rheum Dis. 2011 Jan 5;70(5):805–11.

275. Blom AB, van Lent PLEM, Holthuysen AEM, van der Kraan PM, Roth J, van Rooijen N, et al. Synovial lining macrophages mediate osteophyte formation during experimental osteoarthritis. Osteoarthr Cartil. 2004 Aug 1;12(8):627–35.

276. Gulani V, Calamante F, Shellock FG, Kanal E, Reeder SB. Gadolinium deposition in the brain: summary of evidence and recommendations. The Lancet Neurology. 2017 Jul 1;16(7):564–70.

277. Quantitative Imaging Biomarkers Alliance. Profile: DCE MRI quantification version 1.6 [Internet]. RSNA; 2011 [cited 2019 Feb 4]. Available from: http://qibawiki.rsna.org/images/7/7b/DCEMRIProfile_v1_6-20111213.pdf

278. Heye T, Merkle EM, Reiner CS, Davenport MS, Horvath JJ, Feuerlein S, et al. Reproducibility of dynamic contrast-enhanced MR imaging. Part II. Comparison of intra- and interobserver variability with manual region of interest placement versus semiautomatic lesion segmentation and histogram analysis. Radiology. 2013 Mar;266(3):812–21.

279. Hunter DJ. Pharmacologic therapy for osteoarthritis--the era of disease modification. Nat Rev Rheumatol. 2011 Jan;7(1):13–22.

280. Majumdar S, Newitt D, Jergas M, Gies A, Chiu E, Osman D, et al. Evaluation of technical factors affecting the quantification of trabecular bone structure using magnetic resonance imaging. Bone. 1995 Oct;17(4):417–30.

281. Bennell KL, Creaby MW, Wrigley TV, Hunter DJ. Tibial subchondral trabecular volumetric bone density in medial knee joint osteoarthritis using peripheral quantitative computed tomography technology. Arthritis Rheum. 2008 Sep;58(9):2776–85.

282. Frobell RB, Nevitt MC, Hudelmaier M, Wirth W, Wyman BT, Benichou O, et al. Femorotibial subchondral bone area and regional cartilage thickness: a cross-sectional description in healthy reference cases and various radiographic stages of osteoarthritis in 1,003 knees from the Osteoarthritis Initiative. Arthritis Care Res . 2010 Nov;62(11):1612–23.

283. Wang Y, Wluka AE, Cicuttini FM. The determinants of change in tibial plateau bone area in osteoarthritic knees: a cohort study. Arthritis Res Ther. 2005;7(3):R687–93.

284. MacKay JW, Kapoor G, Driban JB, Lo GH, McAlindon TE, Toms AP, et al. Association of subchondral bone texture on magnetic resonance imaging with radiographic knee osteoarthritis progression: data from the Osteoarthritis Initiative Bone Ancillary Study. Eur Radiol. 2018 Nov;28(11):4687–95.

285. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. Magn Reson Imaging. 2004 Jan;22(1):81–91.

286. Bowes MA, De Souza K, Vincent GR, Conaghan PG. OA may not be as structurally heterogeneous as expected: Shape analysis of all knees from the Osteoarthritis Initiative reveals a consistent pattern of bone shape change over 8 years. Osteoarthr Cartil. 2016 Apr;24, Supplement 1:S254–5.

287. Barr AJ, Dube B, Hensor EMA, Kingsbury SR, Peat G, Bowes MA, et al. The relationship between three-dimensional knee MRI bone shape and total knee replacement-a case control study: data from the Osteoarthritis Initiative. Rheumatology (Oxford). 2016 Sep;55(9):1585–93.

288. Neogi T, Li S, Peloquin C, Misra D, Zhang Y. Effect of bisphosphonates on knee replacement surgery. Ann Rheum Dis. 2018 Jan;77(1):92–7.

289. Vaysbrot EE, Osani MC, Musetti M-C, McAlindon TE, Bannuru RR. Are bisphosphonates efficacious in knee osteoarthritis? A meta-analysis of randomized controlled trials. Osteoarthr Cartil. 2018 Feb;26(2):154–64.

290. Karsdal MA, Byrjalsen I, Alexandersen P, Bihlet A, Andersen JR, Riis BJ, et al. Treatment of symptomatic knee osteoarthritis with oral salmon calcitonin: results from two phase 3 trials. Osteoarthr Cartil. 2015 Apr;23(4):532–43.

291. Brandt KD, Radin EL, Dieppe PA, Putte L van de. Yet more evidence that osteoarthritis is not a cartilage disease. Ann Rheum Dis. 2006 Jan 10;65(10):1261–4.

292. Larkin J, Lohr TA, Elefante L, Shearin J, Matico R, Su J-L, et al. Translational development of an ADAMTS-5 antibody for osteoarthritis disease modification. Osteoarthr Cartil. 2015 Aug;23(8):1254–66.

293. Eckstein F, Wirth W, Guermazi A, Maschek S, Aydemir A. Brief report: intraarticular sprifermin not only increases cartilage thickness, but also reduces cartilage loss: location-independent post hoc analysis using magnetic resonance imaging. Arthritis Rheum. 2015 Nov;67(11):2916–22.

294. Eckstein F, Maschek S, Wirth W, Hudelmaier M, Hitzl W, Wyman B, et al. One year change of knee cartilage morphology in the first release of participants from the Osteoarthritis Initiative progression subcohort: association with sex, body mass index, symptoms and radiographic osteoarthritis status. Ann Rheum Dis. 2009 May;68(5):674–9.

295. Eckstein F, McCulloch CE, Lynch JA, Nevitt M, Kwoh CK, Maschek S, et al. How Do Short-Term Rates of Femorotibial Cartilage Change Compare to Long-term Changes? Four Year Follow-up Data From the Osteoarthritis Initiative. Osteoarthr Cartil. 2012 Nov;20(11):1250–7.

296. Hunter DJ, Bowes MA, Eaton CB, Holmes AP, Mann H, Kwoh CK, et al. Can cartilage loss be detected in knee osteoarthritis (OA) patients with 3-6 months' observation using advanced image analysis of 3T MRI? Osteoarthr Cartil. 2010 May;18(5):677–83.

297. Pedoia V, Su F, Amano K, Li Q, McCulloch CE, Souza RB, et al. Analysis of the articular cartilage T1ρ and T2 relaxation times changes after ACL reconstruction in injured and contralateral knees and relationships with bone shape. J Orthop Res. 2017;35(3):707–17.

298. Joseph GB, McCulloch CE, Nevitt MC, Gersing AS, Schwaiger BJ, Kretzschmar M, et al. Medial femur T2 Z-scores predict the probability of knee structural worsening over 4-8 years: Data from the osteoarthritis initiative. J Magn Reson Imaging. 2017;46(4):1128–36.

299. Chen W, Takahashi A, Han E. 3D Quantitative Imaging of T1rho and T2. In: Proc Intl Soc Mag Reson Med. Montreal; 2011. p. 231.

300. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: A Toolbox for Intensity-Based Medical Image Registration. IEEE Transactions on Medical Imaging. 2010 Jan;29(1):196–205.

301. MacKay JW, Low SBL, Smith TO, Toms AP, McCaskie AW, Gilbert FJ. Systematic review and meta-analysis of the reliability and discriminative validity of cartilage compositional MRI in knee osteoarthritis. Osteoarthr Cartil. 2018 Sep;26(9):1140–52.

302. Bruynesteyn K, Boers M, Kostense P, van der Linden S, van der Heijde D. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. Ann Rheum Dis. 2005 Feb;64(2):179–82.

303. Omoumi P, Babel H, Jolles BM, Favre J. Cartilage can be thicker in advanced osteoarthritic knees: a tridimensional quantitative analysis of cartilage thickness at posterior aspect of femoral condyles. BJR. 2018 Apr 16;91(1087):20170729.

304. Williams TG, Holmes AP, Bowes M, Vincent G, Hutchinson CE, Waterton JC, et al. Measurement and visualisation of focal cartilage thickness change by MRI in a study of knee osteoarthritis using a novel image analysis tool. Br J Radiol. 2010 Nov;83(995):940–8.

305. Eckstein F, Buck R, Wirth W. Location-independent analysis of structural progression of osteoarthritis-Taking it all apart, and putting the puzzle back together makes the difference. Semin Arthritis Rheum. 2017 Feb;46(4):404–10.

306. Cotofana S, Buck R, Wirth W, Roemer F, Duryea J, Nevitt M, et al. Cartilage thickening in early radiographic knee osteoarthritis: a within-person, between-knee comparison. Arthritis Care Res . 2012 Nov;64(11):1681–90.

307. Aspden RM. Osteoarthritis: a problem of growth not decay? Rheumatology. 2008 Oct;47(10):1452–60.

308. Cotofana S, Buck R, Dreher D, Wirth W, Roemer F, Duryea J, et al. Longitudinal (one-year) change in cartilage thickness in knees with early knee osteoarthritis: A within-person between-knee comparison. Arthritis Care Res . 2014 Apr;66(4):636–41.

309. Aigner T, Glückert K, von der Mark K. Activation of fibrillar collagen synthesis and phenotypic modulation of chondrocytes in early human osteoarthritic cartilage lesions. Osteoarthr Cartil. 1997 May 1;5(3):183–9.

310. Goldring MB, Goldring SR. Osteoarthritis. Journal of Cellular Physiology. 2007;213(3):626–34.

311. Burstein D, Gray M, Mosher T, Dardzinski B. Measures of molecular composition and structure in osteoarthritis. Radiol Clin North Am. 2009 Jul;47(4):675–86.

312. Link TM, Li X. Establishing compositional MRI of cartilage as a biomarker for clinical practice. Osteoarthr Cartil. 2018 Sep 1;26(9):1137–9.

313. QIBA CT Volumetry Technical Committee. CT Tumor Volume Change Profile - 2018, Technically Confirmed Profile [Internet]. Quantitative Imaging Biomarkers Alliance; 2018. Available from: http://qibawiki.rsna.org/index.php/Profiles

314. Biomarkers Consortium - PROGRESS OA: Clinical Evaluation and Qualification of Osteoarthritis Biomarkers | FNIH [Internet]. [cited 2019 Jul 4]. Available from: https://fnih.org/what-we-do/biomarkers-consortium/programs/progress-oa

315. Harrington D, Parmigiani G. I-SPY 2 — A Glimpse of the Future of Phase 2 Drug Development? New England Journal of Medicine. 2016 Jul 7;375(1):7–9.